

1 / 154

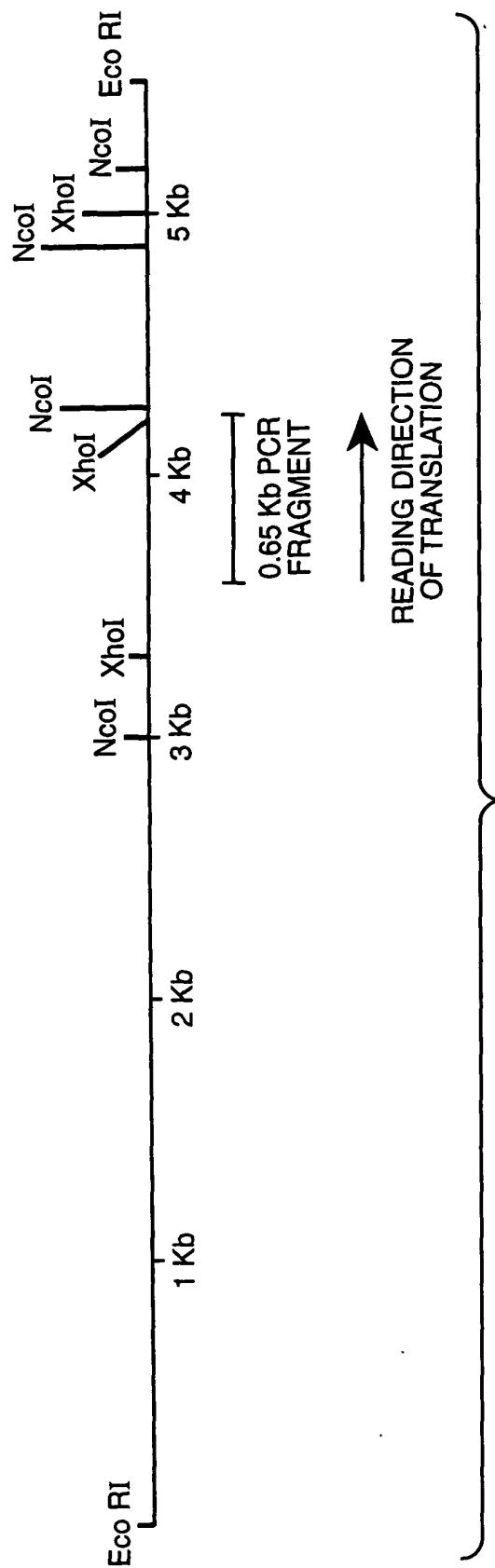


FIG. 1

Nco I
EcoR V
Psp1406 I
CCATGGTGGTGTCACTATCGGCAGTAGTCAGTCTGGCAAACGTTGAGGGTACAGTAGCTGGGAAATCTACGGC 90

Sac I
GGAATATCAAAGTCTTCGAATATCCATATTGGAAAGGACAGAAAGCTCCGGGTAGTTGATAGATGAGCTCCGGTGTATTAAATCGGG 180

BssH II
AGCTGACAGGAGTGAGCGTCATGTAGACCATCTAGTAATGTCAGTCGGCAATTTCGCACATGAACAAGTTGATTTCGGACCCAT 270

Xba I
Bst1107 I
TGT TACATCTCTGGCTACAGCTCGAGATGTGCCCTGCCGAGTAACTTAGAAGCCATGCCAGCTGTGTATAACGACCAAAAGTCAGGG 360

Pvu I
AATATGAAACGATCGTGGATAATTCTTGTGTTTATCCTAAATTAGTCAGGGTTATTAAAGAGATAGATCCCTTCACAAACACT 450

Xba I
CATCCAACGGACTTCTCATACCACTCATGACATAATTCAAAACAGCTCCAGGGCATTAGTTCAACATGAAGCAATTCTCCGAAAC 540
signal sequence

M K O F S A K

FIG. 2A

"Manipulation of the Phenolic Acid Content and Digestibility of Plant Cell Walls by Targeted Expression of Genes Encoding Cell Wall..."

SN# 09/991,209; Dunn-Coleman et al

Docket No. GC648-2

Sheet 3 of 154

Sheet 5 of 154

3 / 154

FIG. 2B

4 / 154

Tth111 I BspM I Acc III
TCCAGGACCAAGTCGAGTCGCTTGTCAAACAGCAGGTTAGCCAGTATCCGGACTATGGCCTGACTGTGACGGCCACAGGTATGCCCTCG 990
V O D O V E S L V K O O V S O Y P D Y A L T V T G H S
introm

TGATTTCTTCAATTAGTGTATAAATACCTCACTCAACTCTACGATAGTCTCGGAGCGTCCCTGGCAGCACTCACTGCCGCCAGCTGTCTG 1080
introm

ApAB I PvuII
L G A S L A A L T A A Q L S
Stu I

BsrG I
CGACATACGACAACATCCGCCCTGTACACCTTGGCGAACCGGGCAATAGGCCCTTCGGCTCGTACATGAACGATGCCCTTCCAAG 1170
A T Y D N I R L Y T F G E P R S G N Q A F A S Y M N D A F O
BspM I Nco I
Xba I
CCTCGAGCCAGATACGACGGCAGTTCCGGGTCACTCATGCCAACGGCATCCAAACCTGCCCGGGTGGAGCAGGGTACGCC 1260
A S S P D T T O Y F R V T H A N D G I P N L P P V E Q G Y A
Sca I

ATGGGGGTAGAGTACTGGAGCGTTGATCCTTACAGCGCCAGAACACATTGTCTGCACCTGGGATGAAGTGCAGTGTGAGGCC 1350
H G G V E Y W S V D P Y S A Q N T F V C T G D E V Q C C E A

FIG. 2C

Ppu10 | BfrB | SnaB | BspLU11 |
 Sph I Bst1107 |

 CGAGTGTACCAAGAAAGATGGATGTCCCTGGAGAGGGCATGCATGTACGTATAACCGAAGGCACACTTTTCGGTAAATCAGGACATGTAAT
 1530

Dra I
BstE II
AAGTTCCTCCATGAATAGATAATGGTTACCCCTACCCATAAGCCTTGGGTTGCCCTTTCTCTTGTGAATATTTAAAGTGA
1620

EcoRV 1710
TGACAGATATCTAAACACCTTATCCGCTTAAACCCATCATAGATTGTCACGTGATAGACCCCTTGAATGATGAGCCAAATGATATCA

Ppu10
Sca I
Dra I
GTCCCCGTTAAATCAAAACCCTTAGGCCACAGTCAGAAATACACCAACCCATTCAAGGTAGTACTAAATATGAAATACAGCCTAAA
1800

| | BfrB I | Bgl II | Ear I | Sap I | Nhe I | Nco I | Eco31 I |
|------|---|--------|-------|-------|-------|-------|---------|
| 1890 | TGGATCGCTATGATCCATAAAGAAGCAACACCTTTCAGATCTCGTTGGCTGGAAAGAGCTAGCTCTACCATGGTCAATTAT | | | | | | |

FIG. 2D

09991209 070202

6 / 154

BamH I
BspLU11 I
BsrG I
GAGTGGAGCCCTTAGCTCGTTAAGCCTAGCTATCCTATAAGGACAACACATGTACATGGGCTTACTTGTAGAGGGTAGGATCCGGG 1980

XbaI
BsrE I
CTTCCTCACATCTCGAGGGAGTTGCTACACGTCGGCCATGTCATAAGCCGGTACTCGACGGTTGTCGTGACCGTGAACCGTGTGACCCAGACCCCTG 2070

XbaI
BsrE I
TGATAGCGTTGAGAAGGCCCTATATTGAAATTCCAAATCTCAGCTTACAGAAGATATGCCCATGGTGGAGGGTTAGTAAACCGATGATGA 2160

Nco I
BsaB I
TGATAGCGTTGAGAAGGCCCTATATTGAAATTCCAAATGCCAAGACCCGCCCTCACCACATGTAAGGCATCCGGCCGGCAC 2250

Eco31 I
Msc I
TCGTGCGAGCATGAGATGAGACCGTGGCCAATCCTGTTCAAATGCCAAGACCCGCCACTCCACGTGGTACTGGCATACTGTGTTCCA 2340

Xcm I
BsrD I
GTTGAATTGCAAAATGCCGAGATCATAAAGCGGCCACACTCCACGTGGTACTGGATGGCTTGCCTGGCCATACTGTGTTCCA 2430

AlwN I
VspI
TTGCCTGGCTGGTACTGGGACGGCAGATTCTGTAGGCAAGGCCAGGGCTCTCTGAGGTAGAAAACACCCCATATTAATCT 2436

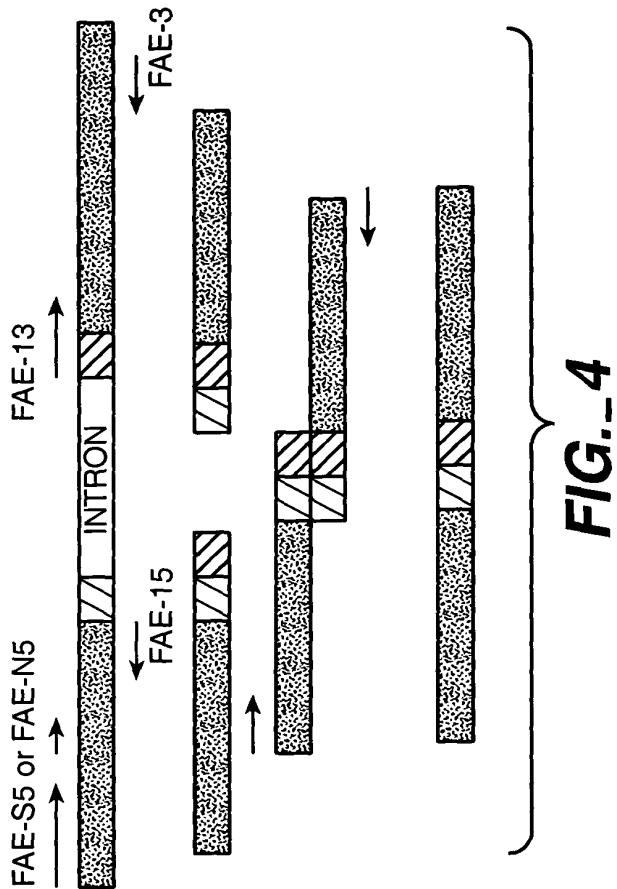
EcoR I
GAATTCA 2436

FIG._2E

CCATGGTGGTGCATATCGGCAGTAGTCTTGCCTGGAAACCGTTACAGTGATCTGGTGGGACATACTT
CGGGGAATCTACGGGGAAATATCAAAGTCTTGGGAGCTTAAATCGGGTGTATTAATCGGGAGCTGACAGGAGTGGAGCTCATGTAACCTCTAGTAATGTCA
GATAGATGAGCTCCGGTGTATTAATCGGGAGCTGAGTGGAGCTGACAGGAGTGGAGCTCATGTAACCTCTAGTAATGTCA
CGGGCAATTTCGCACATGAAACAAGTGAAGCTGGAGTATACTTAAAGCCATGCCAGGGTGTGTTATACGACCAAAAGTCAGGGAAATATGAAACGATCG
TGGGATATTCTGGTTATCCTAAATTAGTCTTCCAGTGGTTATTAAAGATAATGGTTTACAAACACT
TCGGGACTTCTCATACCACTCATGGACATAATTCAAACTCCTAAACTGACATTAACGGATGGGATCTCCTGGGAC
ATTCTCGCCAAACACGTCCTCGCAGTTGGTGAAGTCAGTGCAGGCAACGCCCTAGCAGGCTCTACGCCAGCTGTGCAACATTC
CGGAAGAACCTCTACAGCCGTTAGTCGAAATGGCCACTATCTCCCAAGGCTGCGCAACATTC
CGTGGACATTATCAAGGGAGGAAATTACAATTACCGTCTTCGGCAGCTGGTAGTGTGAAGTACACGGTTGGATATTATGGGATGGGGTCTCCG
GCAAGCAAGAAATAATCACCGTCTTCGGCAGCTGGTAGTGTGAAGTACACGGTTGGGACTATCCGGACTAATCTACCTAC
TCACGCCCTTCGACACCCATCCACAAATGCAACGGCAACAGCAGGTTAGCCAGTTGGGACTATCCGGACTAATCTACCTAC
TCCAGGACCAAGTGGAGTGGCTTGTCAAAACAGCAGGCTGGGACATACGACATAACGAAACATCCGGCTGTACACCTTGGGGAACGGGCA
ACAGGTATGCCCTCGTGATTCTTCATTAAAGTGTATAATACCTACCTACCTACGGGACTATGGGACTATGGGCTGTGACTGTGACGGGCC
TGGCAGCAGCACTGCGCCTTGGCTGATGCCAAACGACGGCATCCAAACCTGGGGGATGGGACTTGGGACTGGGCTGTGACTGTGAGGGCC
GGGCAATTCAAGGGCTTGGCTGATGCCAAACGACGGCATCCAAACCTGGGGGATGGGACTTGGGACTGGGCTGTGACTGTGAGGGCC
GGTCACCTCATGCCAAACGACGGCATCCAAACCTGGGGGATGGGACTTGGGACTGGGCTGTGACTGTGAGGGCC
ACTGGAGCGGTGATCCTTACAGGGCCATCCAAACGACCATTTGGCTGCAACTGGGACTGGGATGAAGTGCAGTGGCTGTGAGGGCC
TCCAGGACAGGGTGTGATAATGGGACAGGACTTATTGGGATGGGACTGGGCTGTGACTGGGCTGTGACTGTGAGGGCC
AGGGGGACAGGGTGTGATAATGGGACAGGACTTATTGGGATGGGACTGGGCTGTGACTGGGCTGTGACTGTGAGGGCC
TCATTTCAGCCTCCGGTAAATCAGGACATGTATAAGTTCCTCCATGAAATAGGATATGGTTACCCCTCACCATAAAGCCTT
ACACTTTTCCGGTAAATCAGGACATGTATAAGTTCCTCCGGTAAATATGGTTACCCCTCACCATAAAGCCTT
GAGGTTGGCTTCTCTTGGATTGGATAATTAAAGTAGATGACAGATAATCTCTAAACACCTTATCCGGCT
TAACCCATCATAGATTGGTGTACCGTGTAGGACCCCTTGAATGAGGAAATGGGAAATGGTATCAGTCCCCTGGGAAGAGCTAGCTCTAC
AACCCCTTTCAGCCTAGCAGTCAAGTCAAGAATAACACCAACCCATTCTAAGGTAGTACTAAATGAAATACAGCCTAA
TGCACTCGCTTATGATGCCATAAAAGCAACACCTTCAGATCTCGTTAGCTTAAAGGACAACACCTTCA
CATGGTCTCAATTATGAGGTGGAGCTTGTCTCGTTAGGATCCGGGCTTCTCACATCTCGAGGAGTTGTTGATAAGCCTGTTGAGAAGGGCT
GGCTTACTTGTAGAGGGTAGGATCCGGGCTTCTCACATCTCGAGGAGTTGTTGATAAGCCTGTTGAGAAGGGCT
TAAGCCGGTACTCCGACGTTGTGACCGTGAACCCAGACCCAGACCCCTGGGAGGGTGTAGTAAGGCAACCTTCA
TTTCCCAATTCTCAGGCTTACGAAGATAATGCCATTGGGGCTTGTAGTAAGGCAACCTTCA
GATGAGACCCGGTGGCCAATTCTGTCAATTGCCAAGACCCGGCTTCTCACACATGTAAGGCAACCTTCA
GTGAAATTGTGCAAATGCCGAGATCATAAAAGGGCCACACTTCCACGTGGTACTGGGATGGGTT
ATACTGGTGTGTTCCATTGGGTGGTGTGTTACTGGGACGGCAGATTCTGTAGGCAAGGGCAAGGGCT
TCTGAGGGTAAACACCCATTTAAATCTGAATT

FIG.-3

8 / 154



FAE-I3 CCGGCCACGCCCTGGCCCCTGGCGGCACTC
 FAE-I5 GGCCCGAGGGAGTGGCCGGTAGTCC

i

Y A L T V T G H S L G A S L A A L
 GGACTACGGCTGACCGTGACCGCCACTCCCTGGGGCC
 CCGGCCACGCCCTGGGCCCCTGGCGGCACTC
 Y A L T V T G H A L G A S L A A L

complement, FAE-15
 FAE-13

FIG. 5

Vector Construction

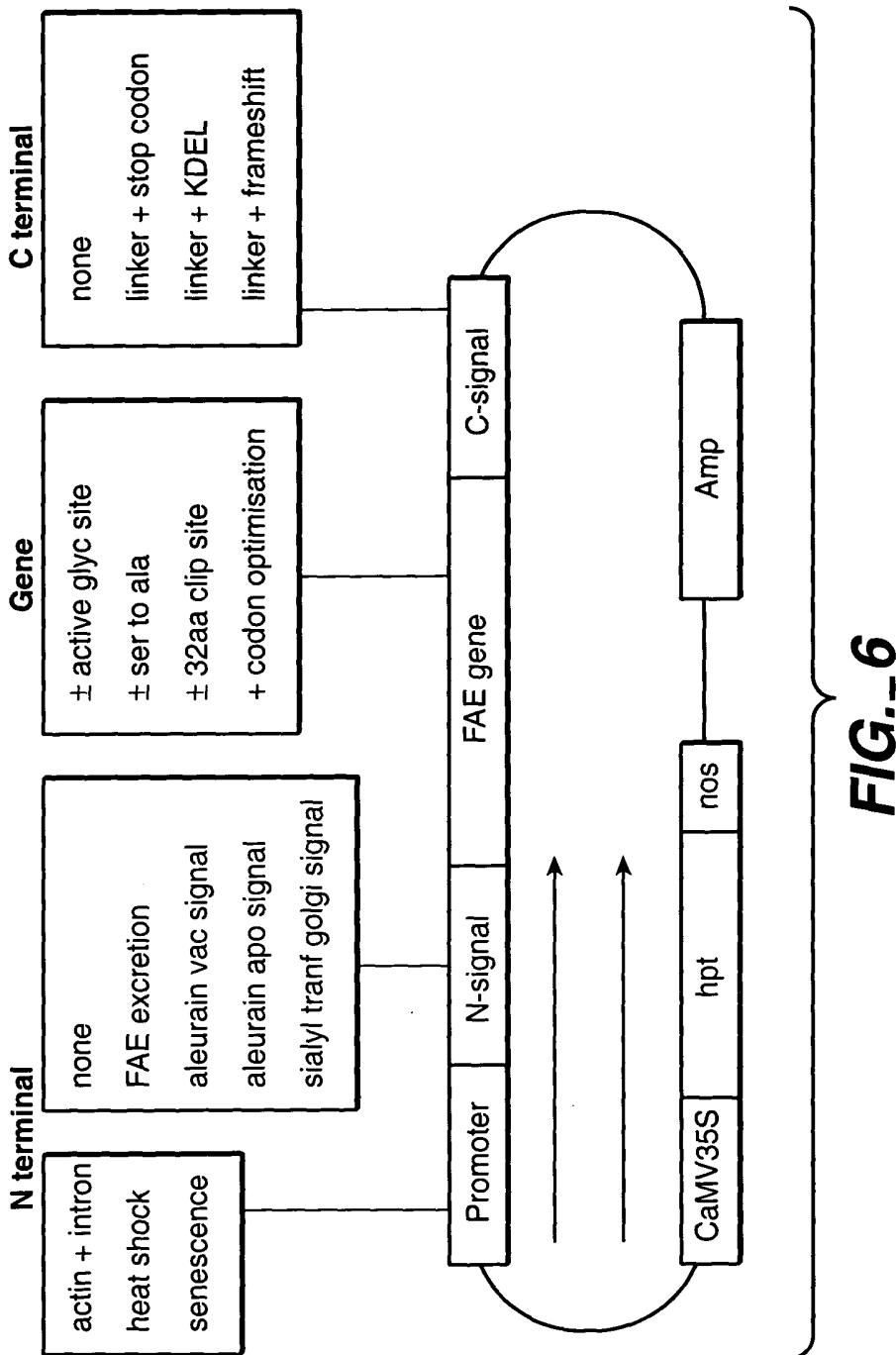


FIG._6

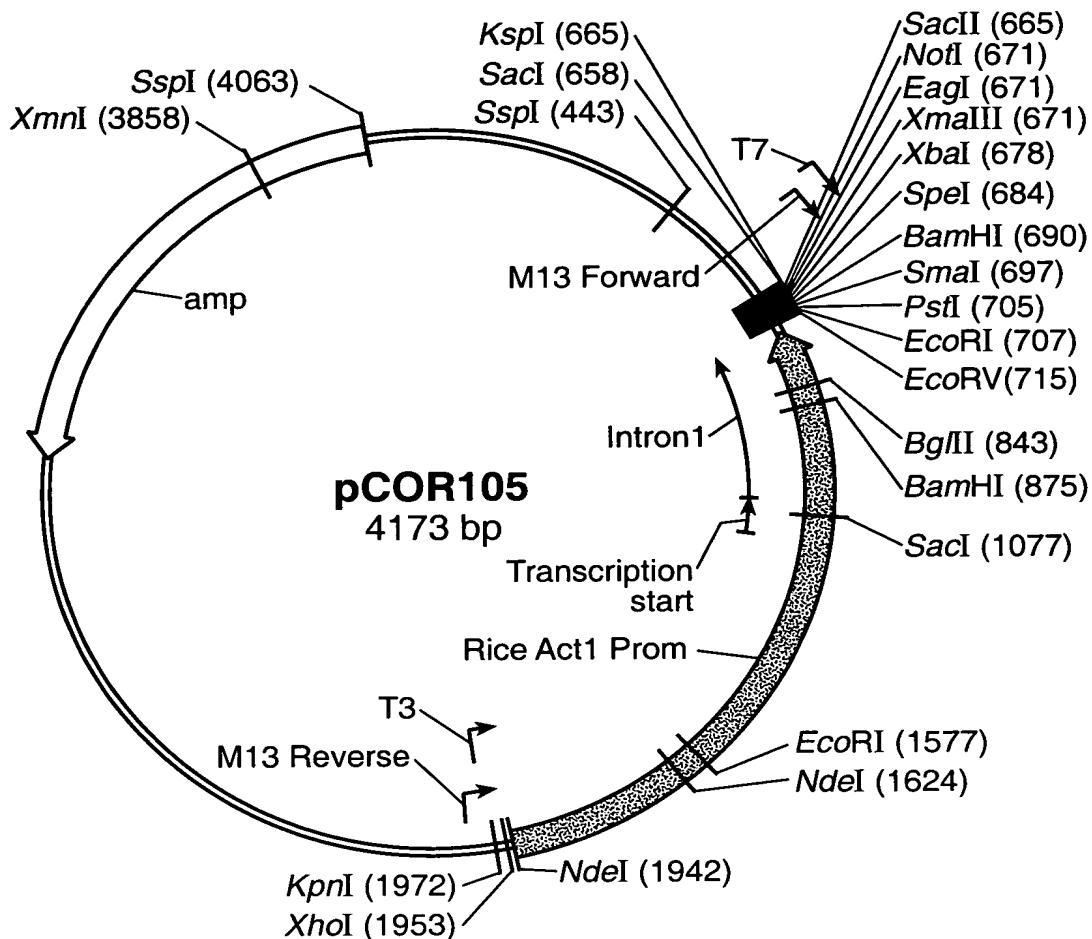


FIG._7

11 / 154

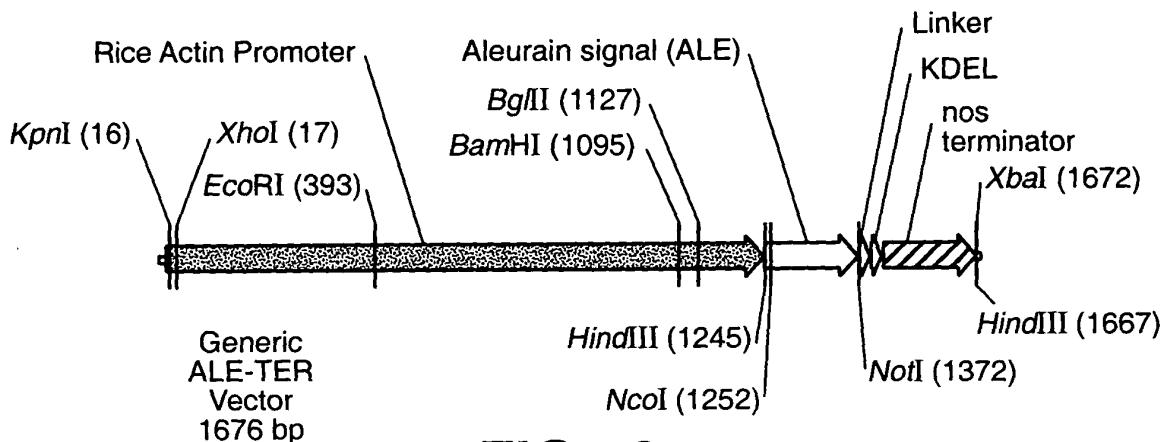


FIG._8

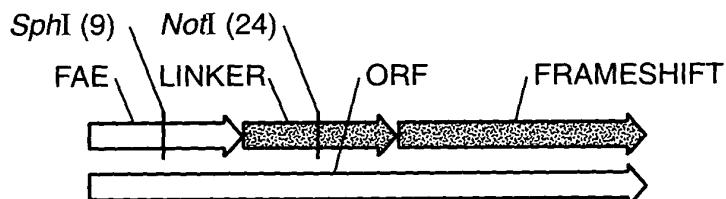
KDEL-COOH ER retention sequence

NotI

1 A A A K P L K D E L *
1 GCGGCCGCGA AACCACTGAA GGATGAGCTG TAA

FIG._9

FAE-LINKER-FRAMESHIFT Structure and Sequence



FAE-LINKER-FRAMESHIFT
48 bp

+1 G A C T W P V A A A E T T E G
SphI NotI

1 GGCATGCA CCTGGCCGGT CGCGCCGCG GAAACCACTG AAGGATGA
CCCGTACGT GGACCGGCCA GCGCCGGCGC CTTTGGTGAC TTCTACT

FIG._10

| Initial Vectors | HS | Original Actin + hyg | Plant Transformation Cassettes | | | Actin (+hyg) | H.S. | See1 |
|-----------------|-----|----------------------|--------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| | | | Target | Actin | H.S. | | | |
| TP11.1 | TT3 | TR9.4 | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| TT5 | - | TR5.5 | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| UA4.4 | - | - | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| TP8.5 | - | - | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| TP3.1 | - | TR8 (glycos) | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| TU4 | - | - | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| TU5 | - | - | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| UG | - | - | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| TP5.1 | TT2 | TR6.1 | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| TP4 | - | TR2 | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| TP3.1 | - | - | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| TP3.1 | - | - | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| TP3.1 | - | - | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg | Original Actin + hyg |
| | | | | | | | | |

* - Modified Actin Promoter (Kpn1-EcoR1 Deletion and Restored NCO Site)

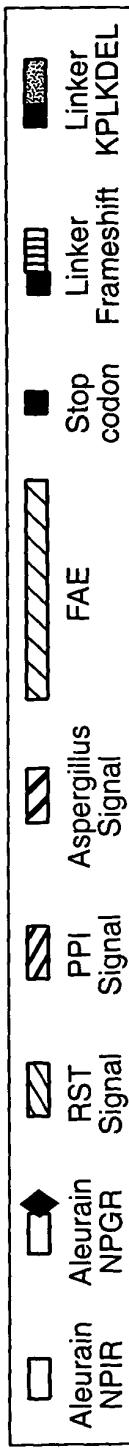


FIG.-11

Vectors

Original Actin promoter in pCOR105

| | Target | Signal sequences | Vectors |
|-------|---------------|---|---|
| (i) | APO | - aleurain-NPGR-FAE - aleurain-delNPIR -FAE | pUH6, pTT5, TT5.5, pTT5.1 |
| (ii) | ER | - aleurain-NPGR-FAE-linker-KDEL - aleurain-delNPIR-FAE-linker-KDEL | pUH7, pUA4.4, pTU5, pUH8, pUG4, pUH9, |
| (iii) | VAC | - aleurain-NPIR-FAE | pTP11.1, pTR9.4, pUH4, pUK3, |
| (iv) | ER/VAC | - aleurain-NPIR-FAE-linker-KDEL | pTU4, pUH3, |
| (v) | VAC | - aleurain-NPIR-FAE-linker-frameshift | pUA1K3, pTP3.1, pUC5.11 |
| (vi) | VAC | - aleurain-NPIR-FAE-linker-stop | pTP8.5, pUH5 |
| (vii) | ER | - Aspergillus signal -FAE-KDEL | pTP5.1, pTP6.1, pUF1, |

Modified actin promoter (Kpn1-EcoR1 deletion and restored NCO site)

| | | | |
|-------|-------|---------------------------------------|--------|
| (i) | VAC | - aleurain-NPIR-FAE-linker-frameshift | pJ06.3 |
| (ii) | GOLGI | - RST-FAE-linker-frameshift | pJQ3.2 |
| (iii) | APO | - PPI-FAE-linker-frameshift | pJQ4.9 |

Heat-shock promoter

| | | | |
|-------|--------|---|------------------------------------|
| (i) | APO | - aleurain-NPGR-FAE - aleurain-delNPIR-FAE - Aspergillus signal-FAE | pUH12 pUH13 pTP4a2, pTR2.22, |
| (ii) | ER | - aleurain-NPGR-FAE-linker-KDEL - aleurain-delNPIR-FAE-linker-KDEL | pUH10 pUH11 |
| (iii) | VAC | - aleurain-NPIR -FAE | pUK3,pTT3 |
| (iv) | ER/VAC | - aleurain-NPIR-FAE-linker-KDEL | pUK2 |
| (v) | VAC | - aleurain-NPIR-FAE-linker-frameshift | pUC5.11, pHOX3 |
| (vi) | VAC | - aleurain-NPIR-FAE-linker-stop | pUK6 |
| (vii) | ER | - Aspergillus signal -FAE-KDEL | pUK1, pTT2 |

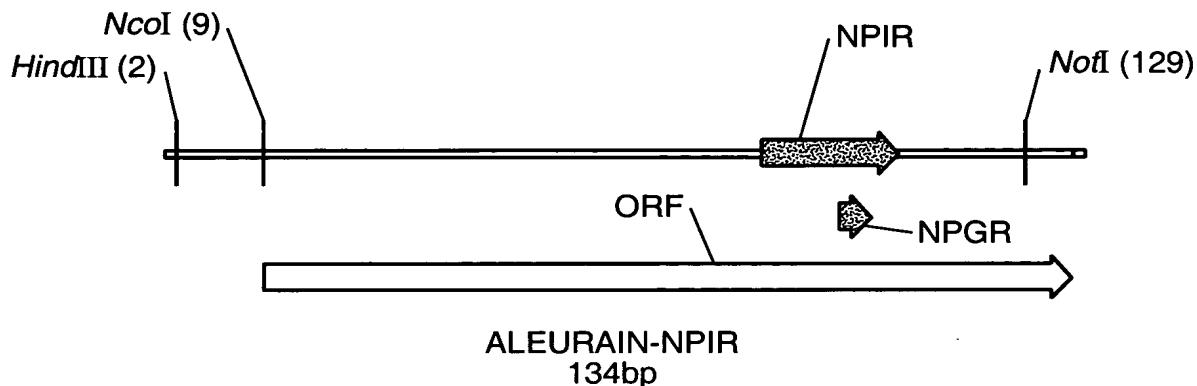
Senescence promoter

| | | | |
|------|-----|----------------------------------|--------|
| (i) | APO | - See1-PPI-FAE-linker-frameshift | pJQ5.2 |
| (ii) | VAC | - See1-aleurain-deleted NPIR-FAE | pUB8.1 |

FIG._12

**ALEURAIN-NPIR (Vacuolar) and NPGR (Apoplast)
Structure and Sequence**

NPIR Underline
NPGR Bold



+1 M A H A R V L L L A L A V L A T A A V A
HindIII NcoI
~~~~~ ~~~~~~

1 AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCCG  
TTCGAATGGT ACCGGGTGCG GGCGCAGGAG GAGGACCGCG AGCGGCACGA CCGGTGCCGG CGGCAGCGGC

+1 V A S S S S F A D S N P I R P V T D R A A  
NotI  
~~~~~

71 TCGCCTCCTC CTCCTCCTTC GCCGACTCCA ACCCGATCCG GCCCGTCACC GACCGCGCGG CCGC
AGCGGAGGAG GAGGAGGAAG CGGCTGAGGT TGGGCTAGGC CGGGCAGTGG CTGGCGCGCC GGCG

FIG._ 13

RAT SIALYL TRANSFERASE Golgi signal sequence

HindIII

~~~~~

1    M    I    H    T    N    L    K    K    F    S    L    F    I    L    V    F    L    L    F    A

1    AAGCTTACCA TGATCCACAC CAACTCAA AAGAAAGTCT CCTCTCTTCAT CCTCGTCTTC CTCCTCTTCG

71    .    V    I    C    V    W    K    K    G    S    D    Y    E    A    L    T    L    Q    A    K    E    F    Q    M

71    CCGTGATCTG CGTGTCAG AAGGCTCCG ACTACGAGG CCTCACCCCTC CAAGCCAAGG AGTTCCAAT

Not I

~~~~~

141 . A A GGGGGCGC

FIG._ 14

POTATO PROTEASE INHIBITOR III Apoplast signal sequence

HindIII

~~~~~

1    M    X    V    H    K    E    V    N    F    V    A    Y    L    L    I    V    L    G    L    L    L

1    AAGCTTACMA TGGMCGTGCA CAAGGAGGTS AACTTCGTSG CCTAACCTCCT GATCGTSCTC

GGCCTCCCT

NcoI

~~~~~

1 . L V S A M E H V D A K A C T X E C G N L

1 G F G .

71 TGCTCGTSTC CGCCATGGAG CACGTGGACG CCAAGGCTG CACCCKGAG TGCGGCAACC

TCGGCTTCGG

Not I

~~~~~

141    .    I    C    P    A    A    A    CATCTGCCCG GGGCCGCC

**FIG.\_ 15**

16 / 154

Targeting Expression of gfp to Different Cell Compartments

Actin Promoter Targeting Vectors

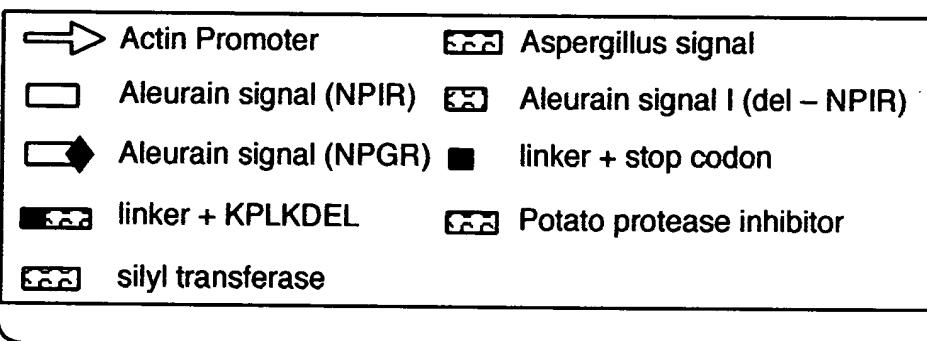
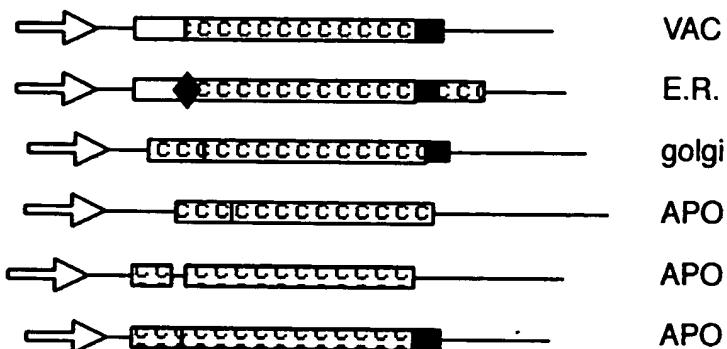


FIG.\_ 16A

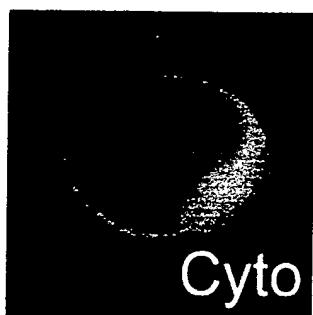


FIG.\_ 16B

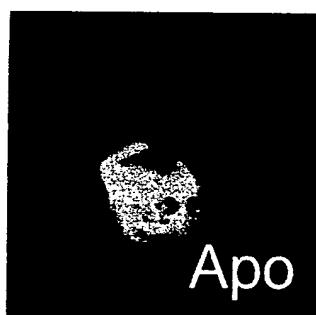


FIG.\_ 16C

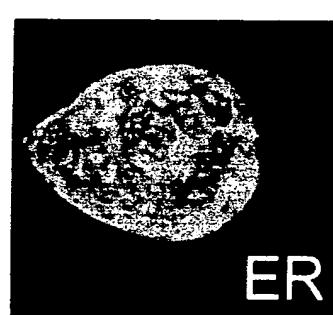


FIG.\_ 16D



FIG.\_ 16E

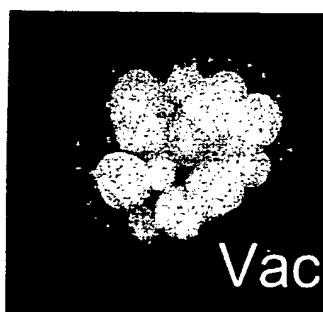


FIG.\_ 16F

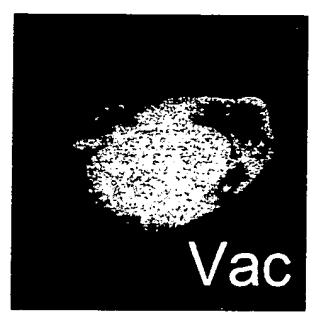
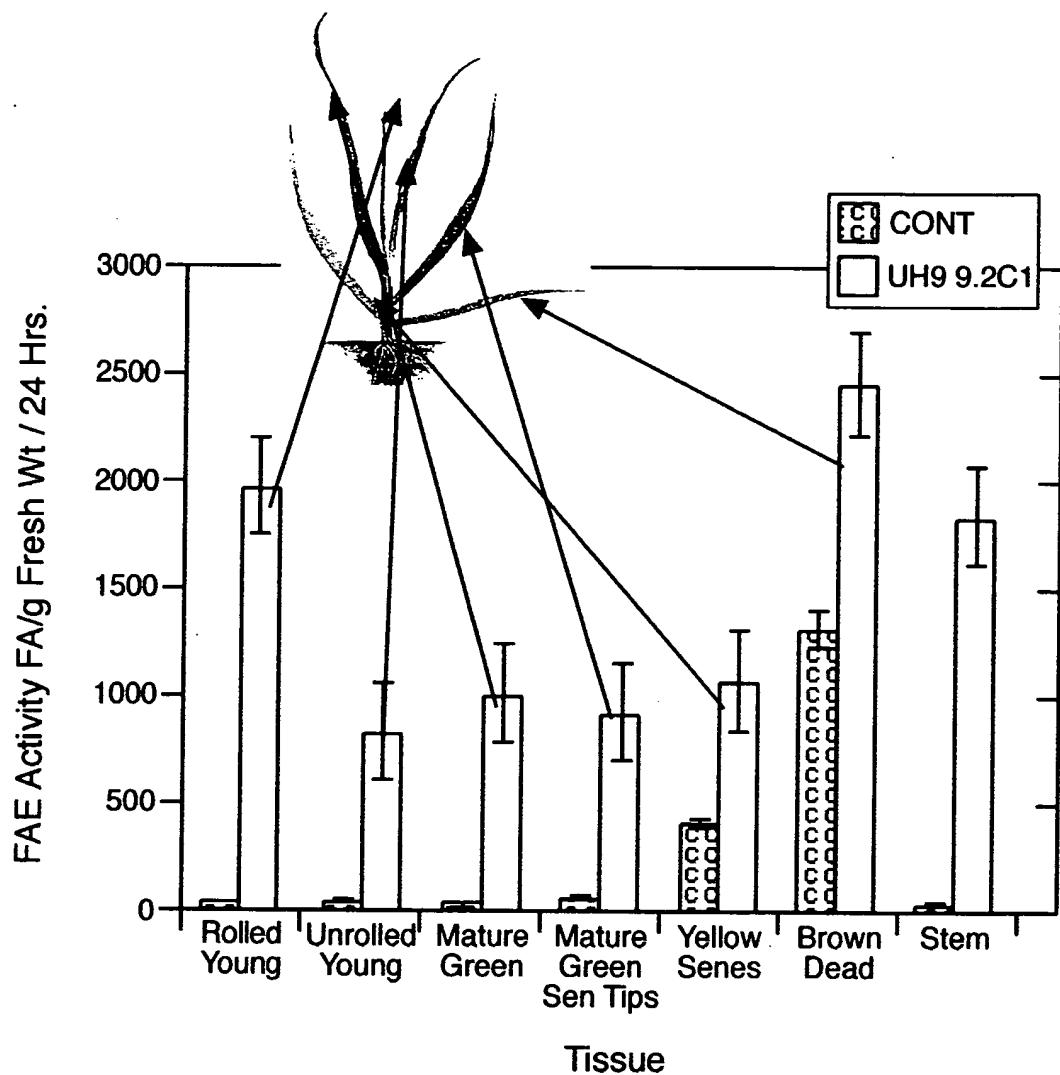


FIG.\_ 16G

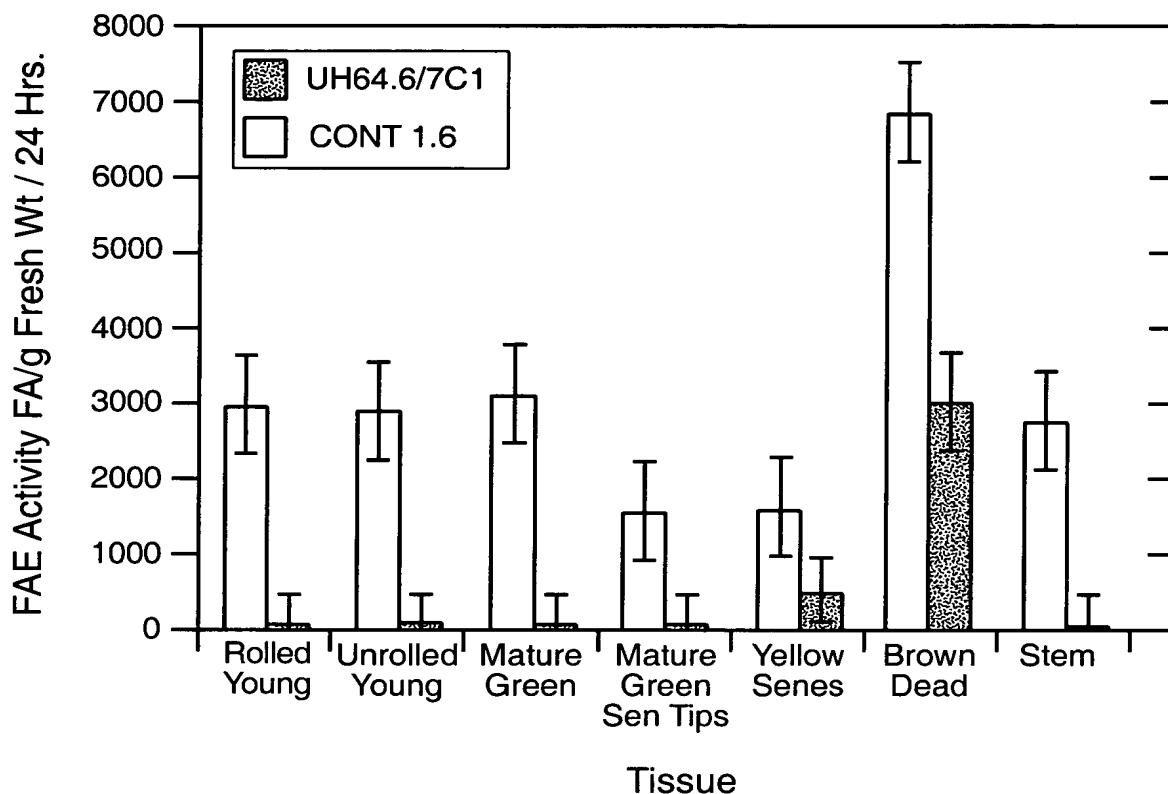
**FAE Activity in Transgenic *Festuca arundinacea* Leaves  
of Different Ages Under ER and APO Targeting Sequence**



**FIG.\_ 17A**

18 / 154

**FAE Activity in Transgenic *F. stucaria* arundinacea L.aves  
of Different Ages Under ER and APO Targeting Sequence**



**FIG.- 17B**

Fae Activity In Leaves of Primary Transformants of  
*Festuca Arundinacea* Under Vac Targeting Sequence

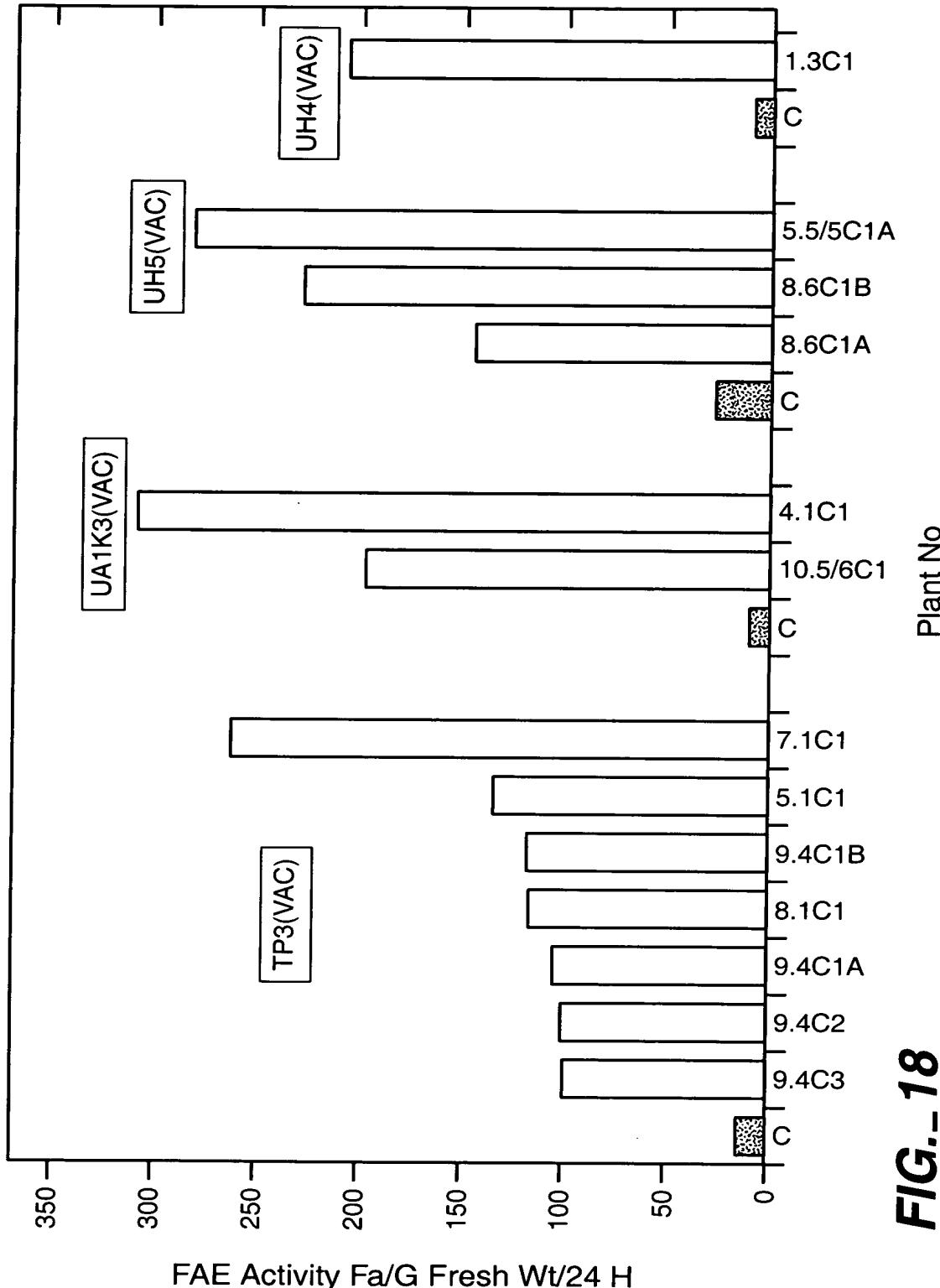
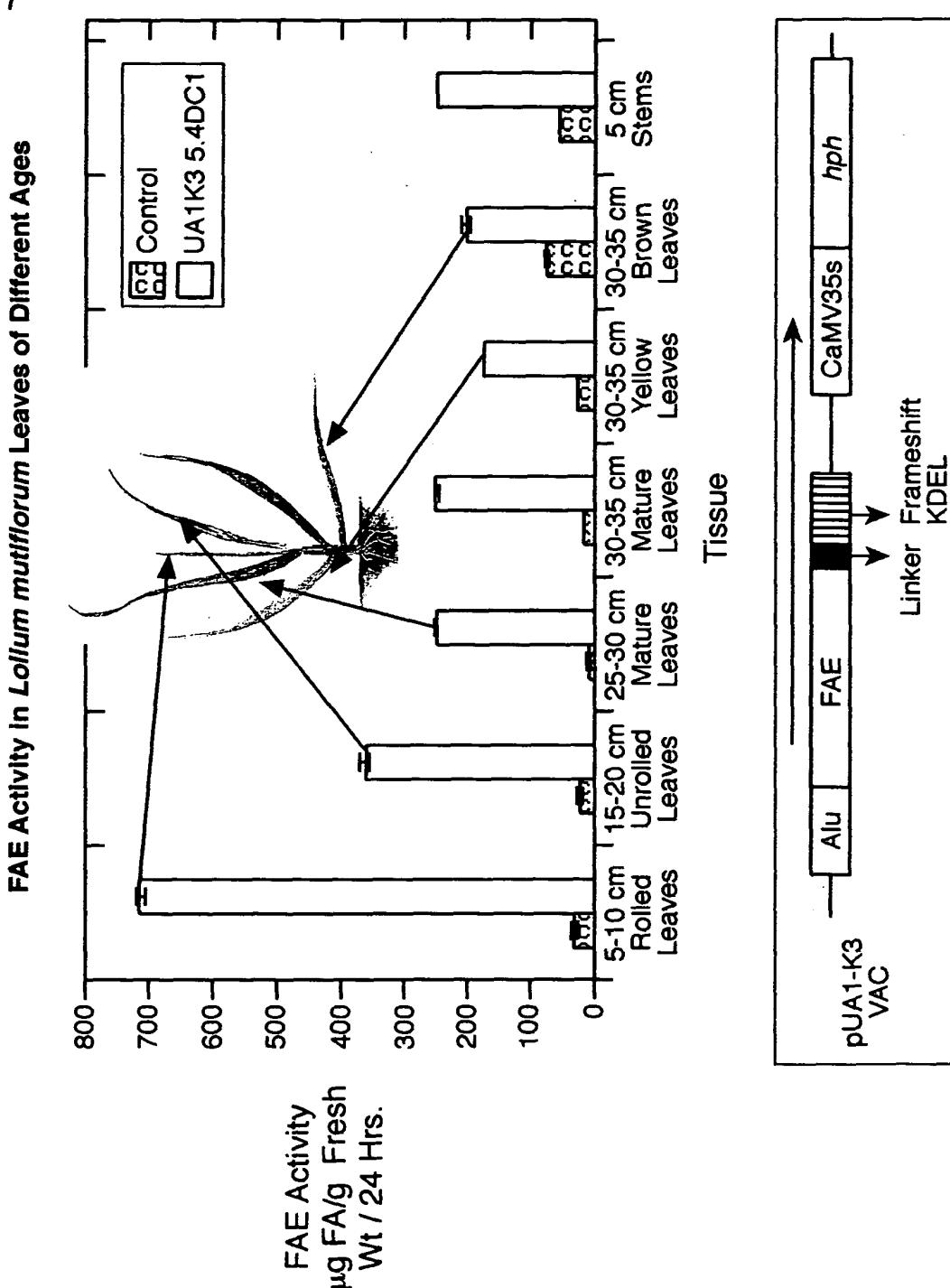


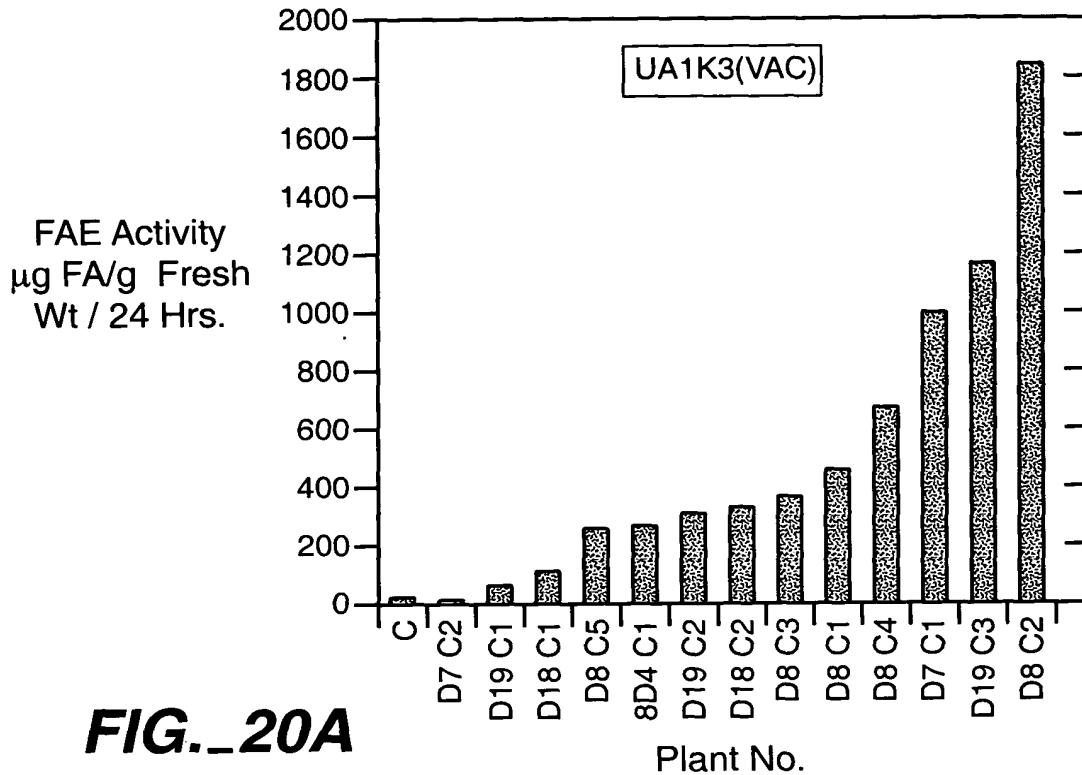
FIG. 18

FIG. - 19



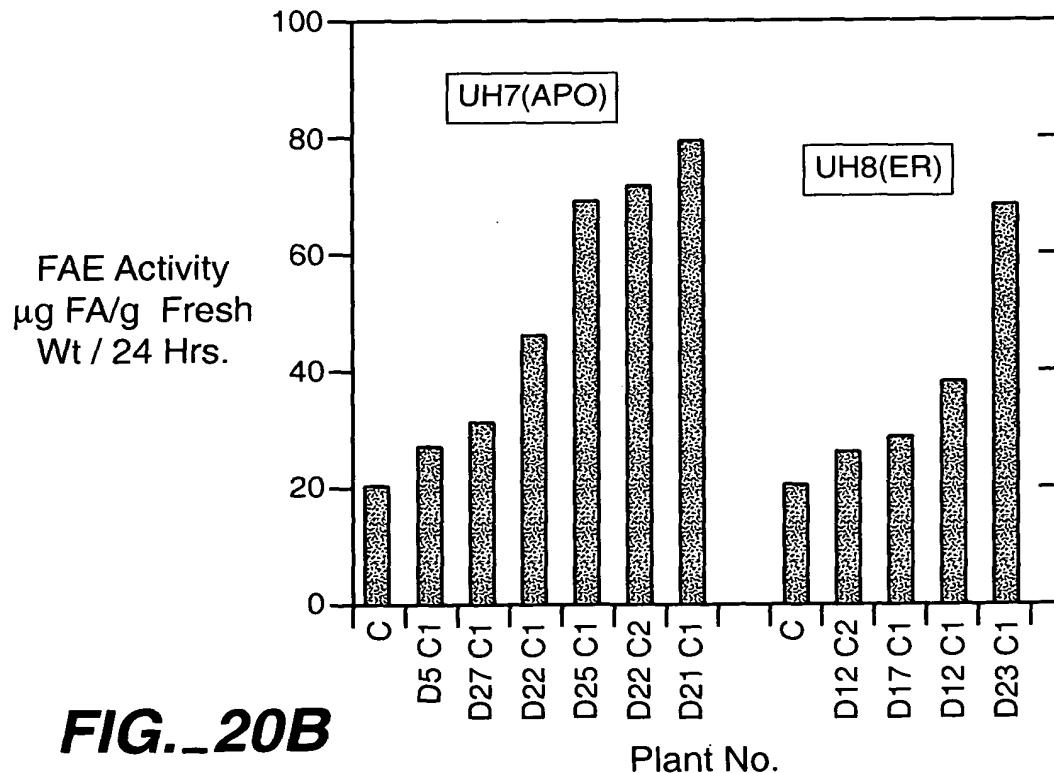
21 / 154

**FAE Activity in Leaves of Primary Transformants of *Lolium multiflorum* Under VAC APO and ER Targeting Sequence**



**FIG.\_20A**

**FAE Activity in Leaves of Primary Transformants of *Lolium multiflorum* Under VAC APO and ER Targeting Sequence**



**FIG.\_20B**

+

Levels of Esterified Monomeric and Dimeric Hydroxycinnamic acids in  
*Festuca Arundinacea* Plants Expressing FAE Under VAC Targeting Sequence

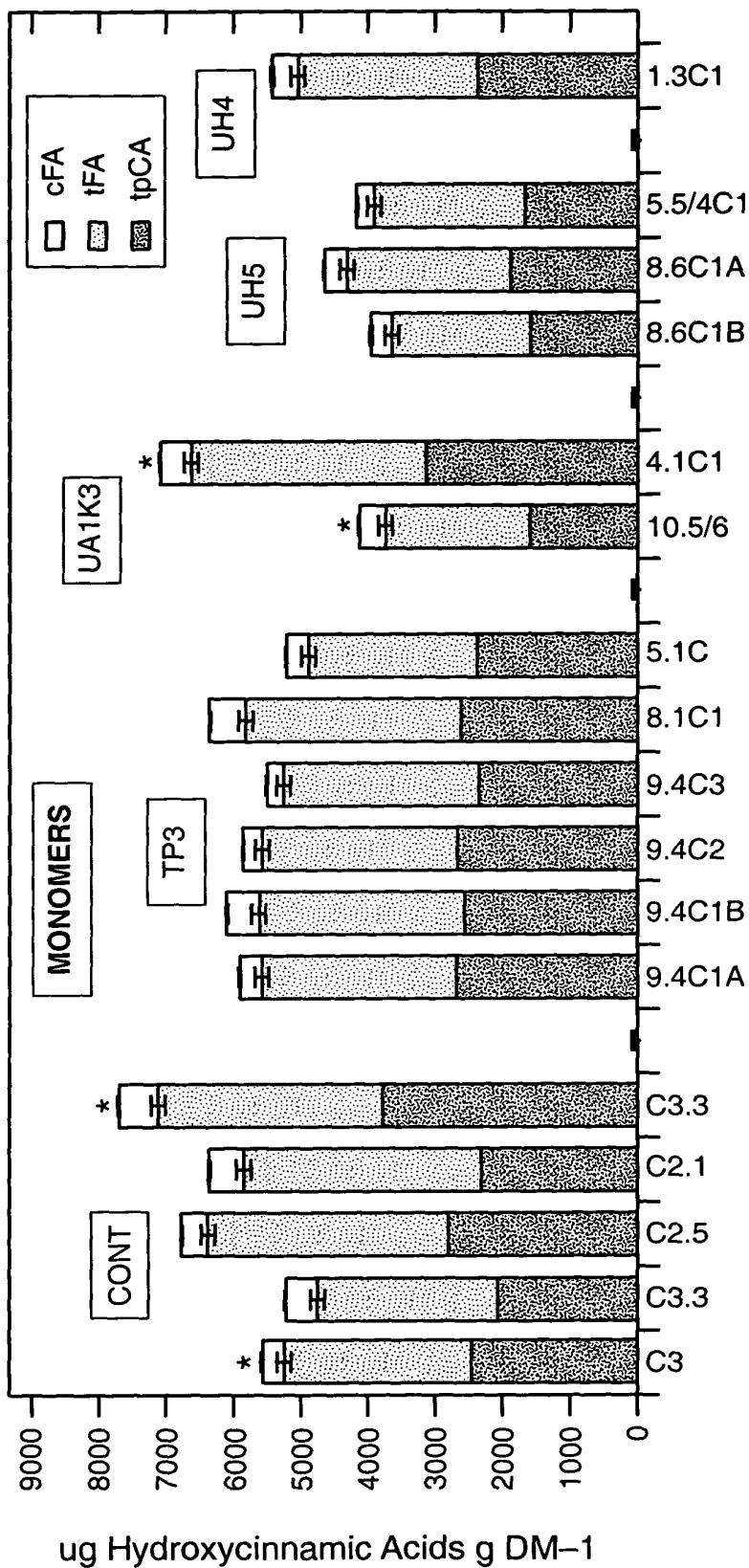
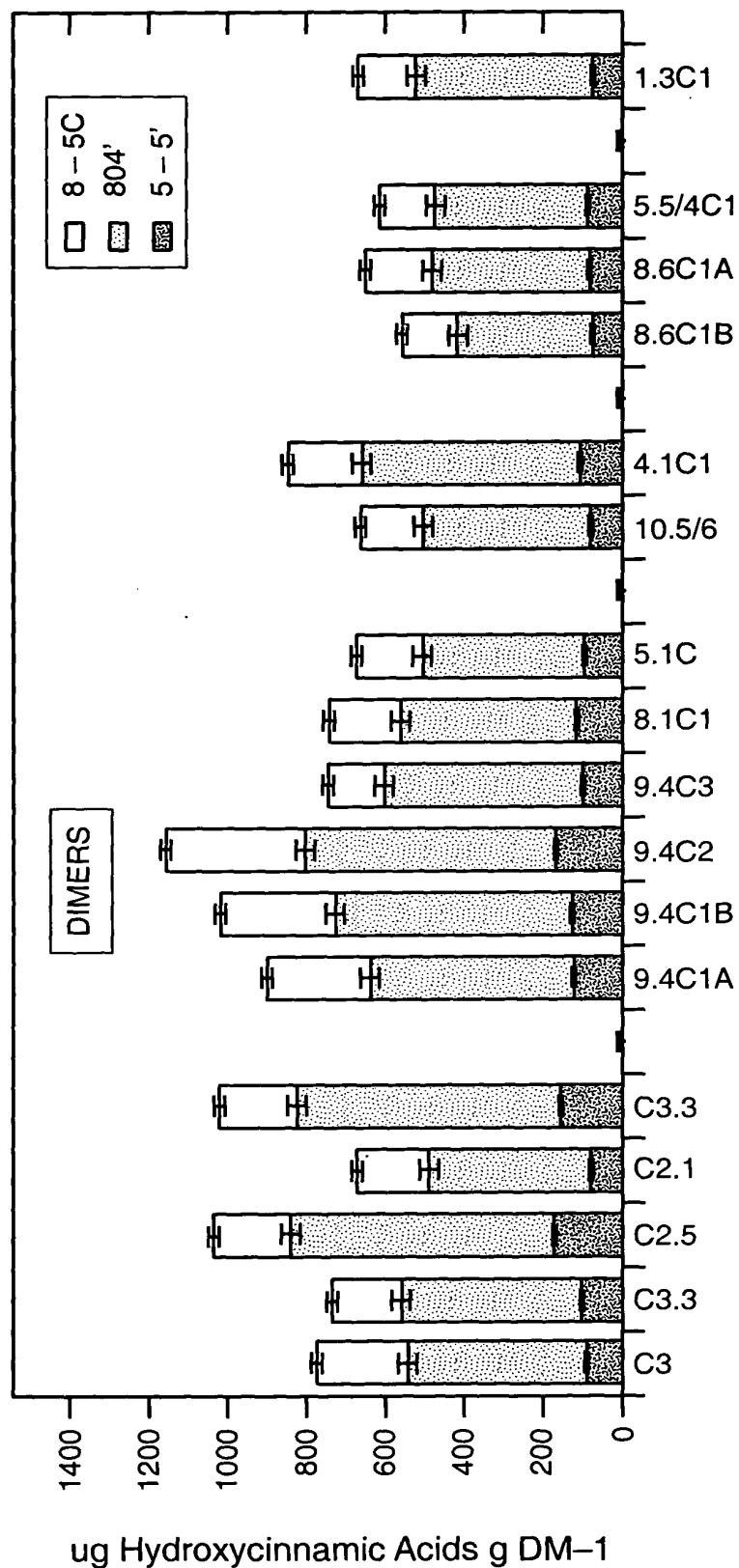
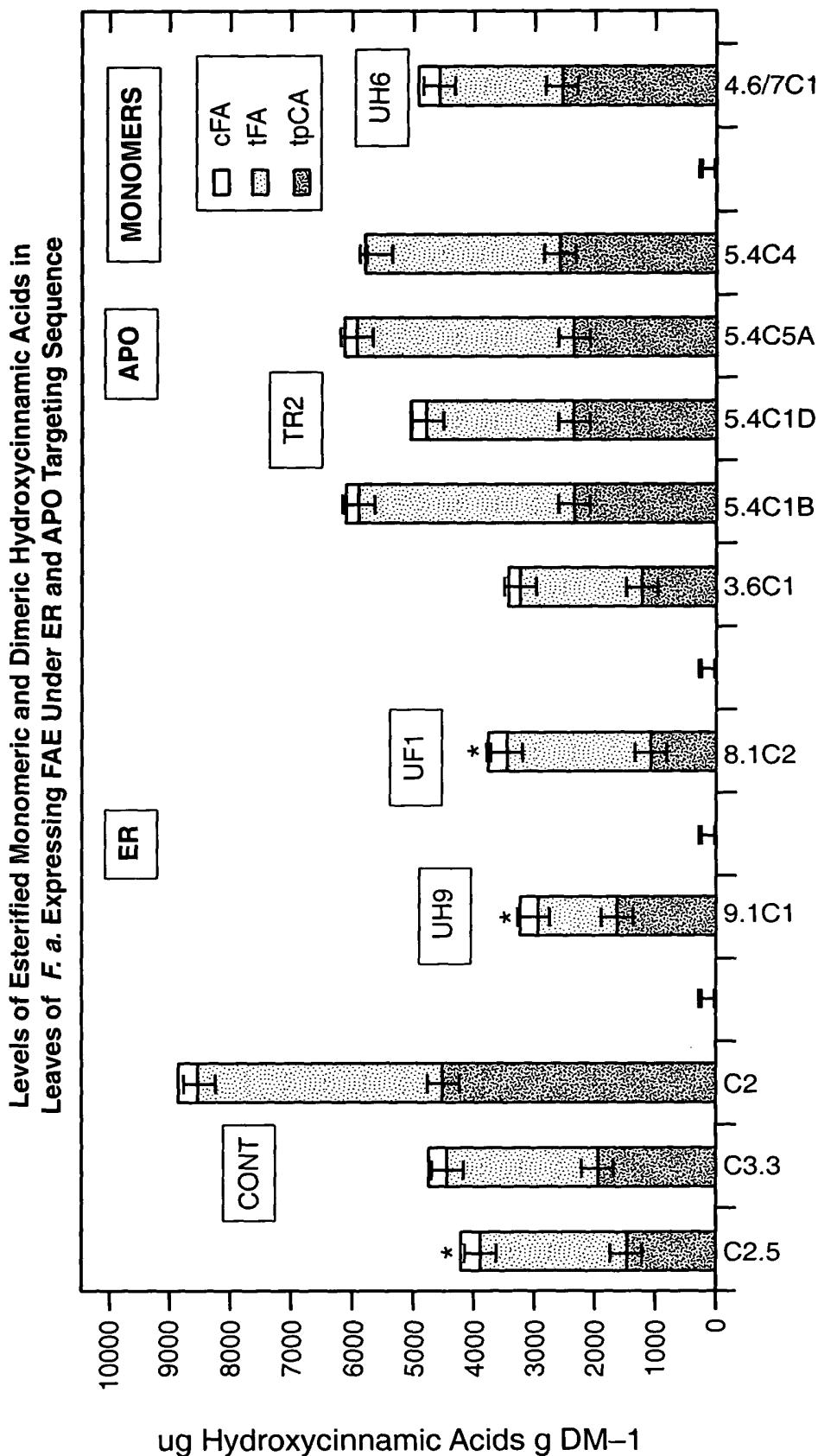


FIG.-21A

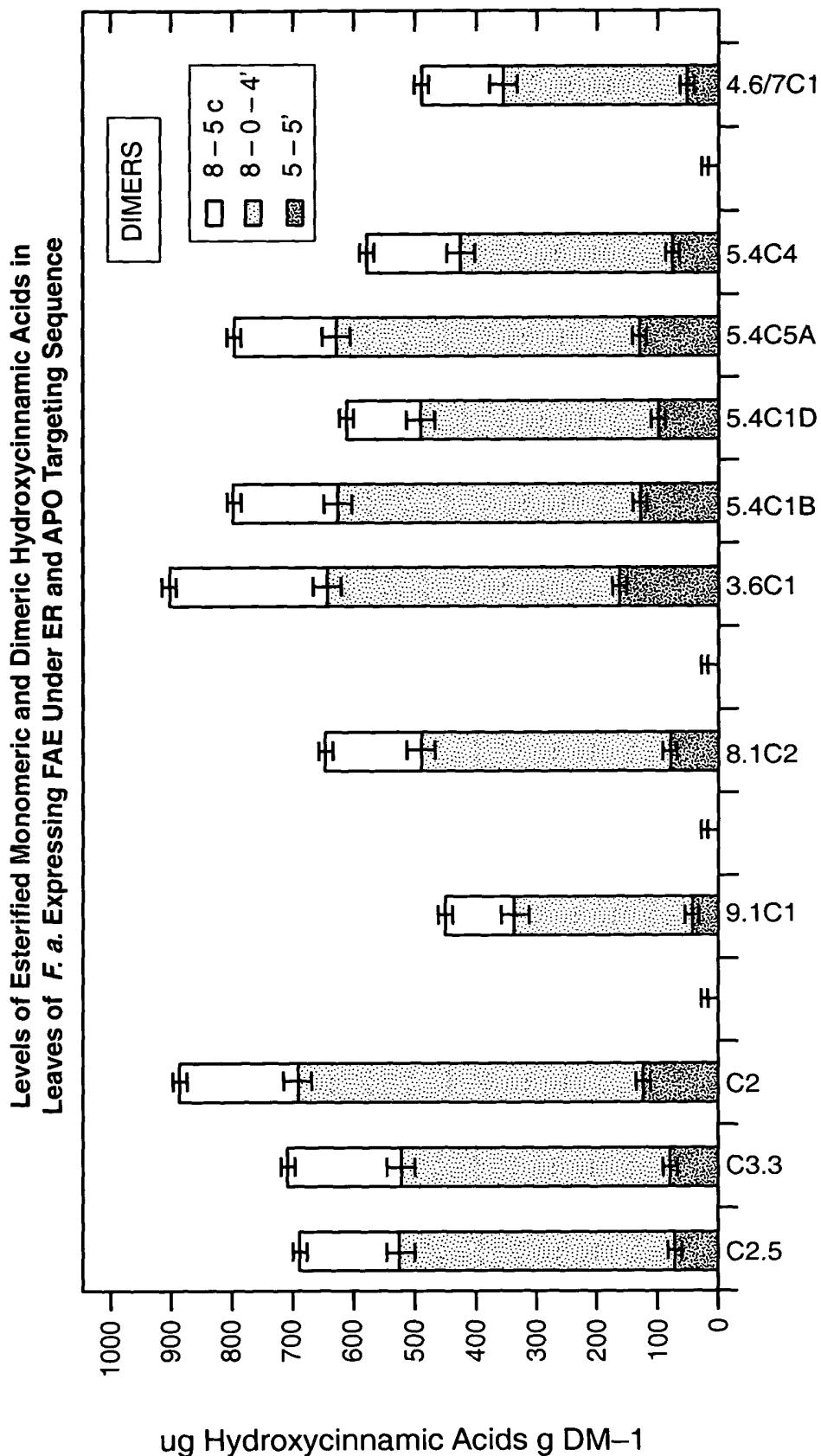
FIG. 21B  
Levels of Esterified Monomeric and Dimeric Hydroxycinnamic acids in  
*Festuca Arundinacea* Plants Expressing FAE Under VAC Targeting Sequence

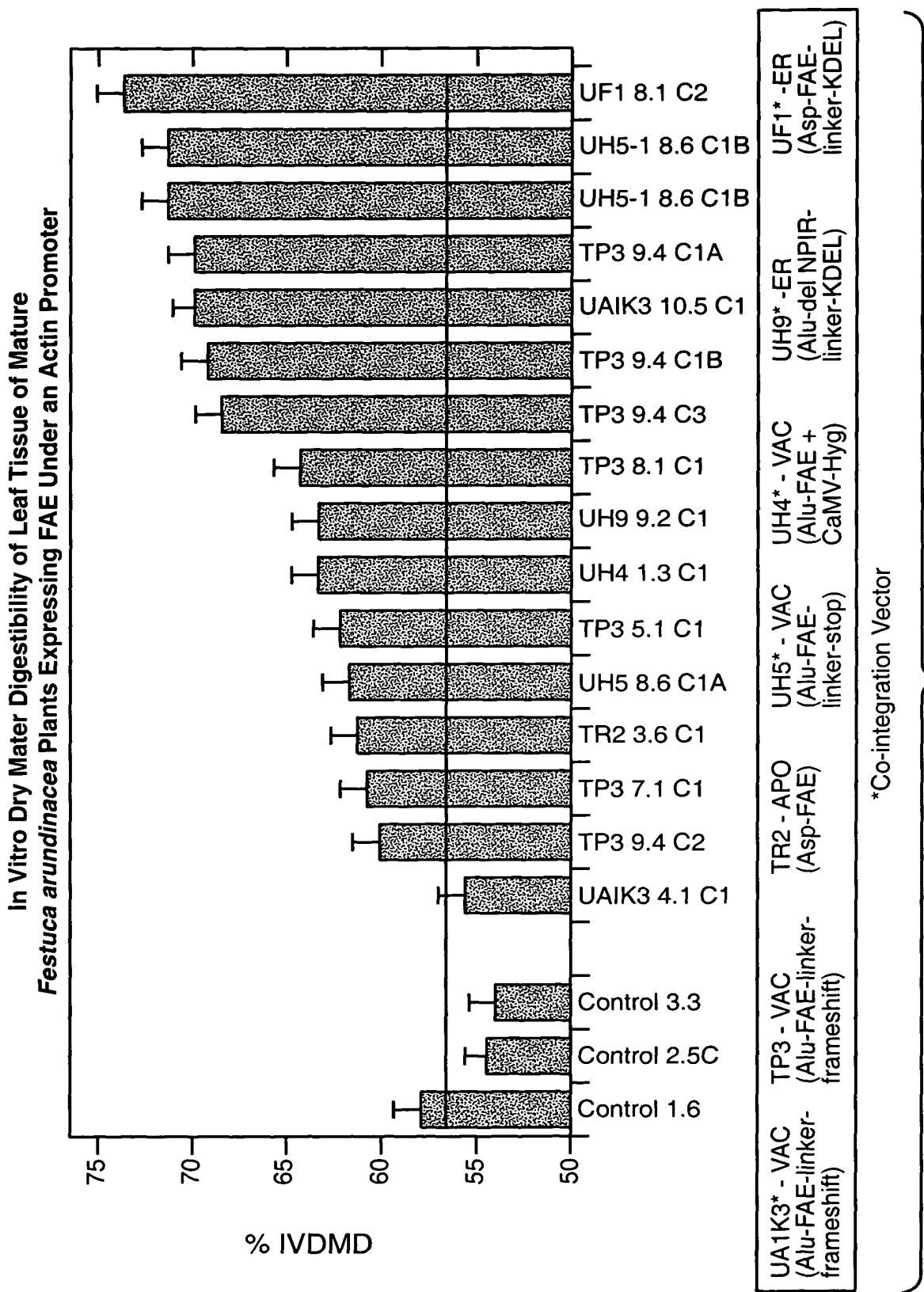




**FIG. 22A**

25 / 154





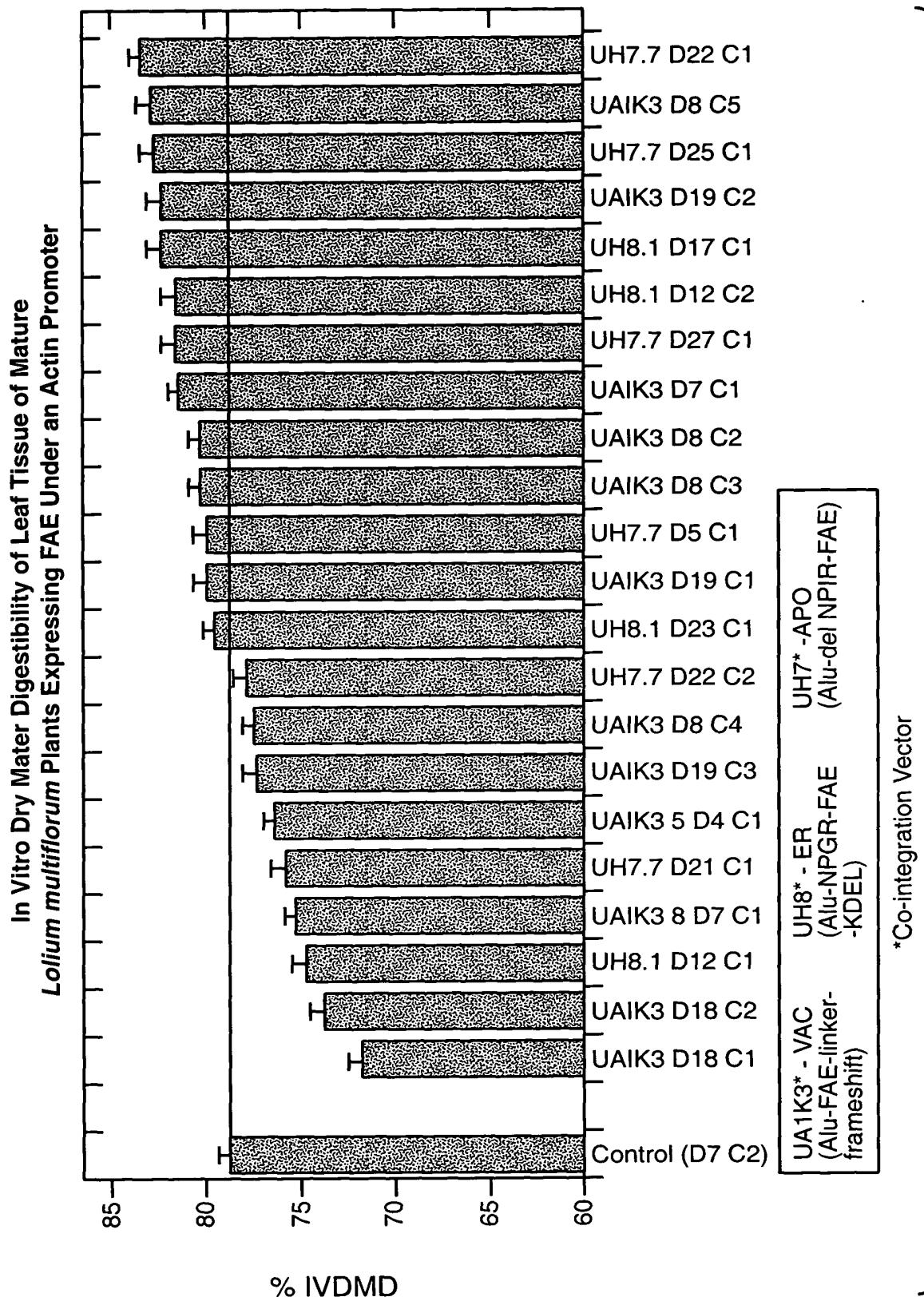


FIG.\_24

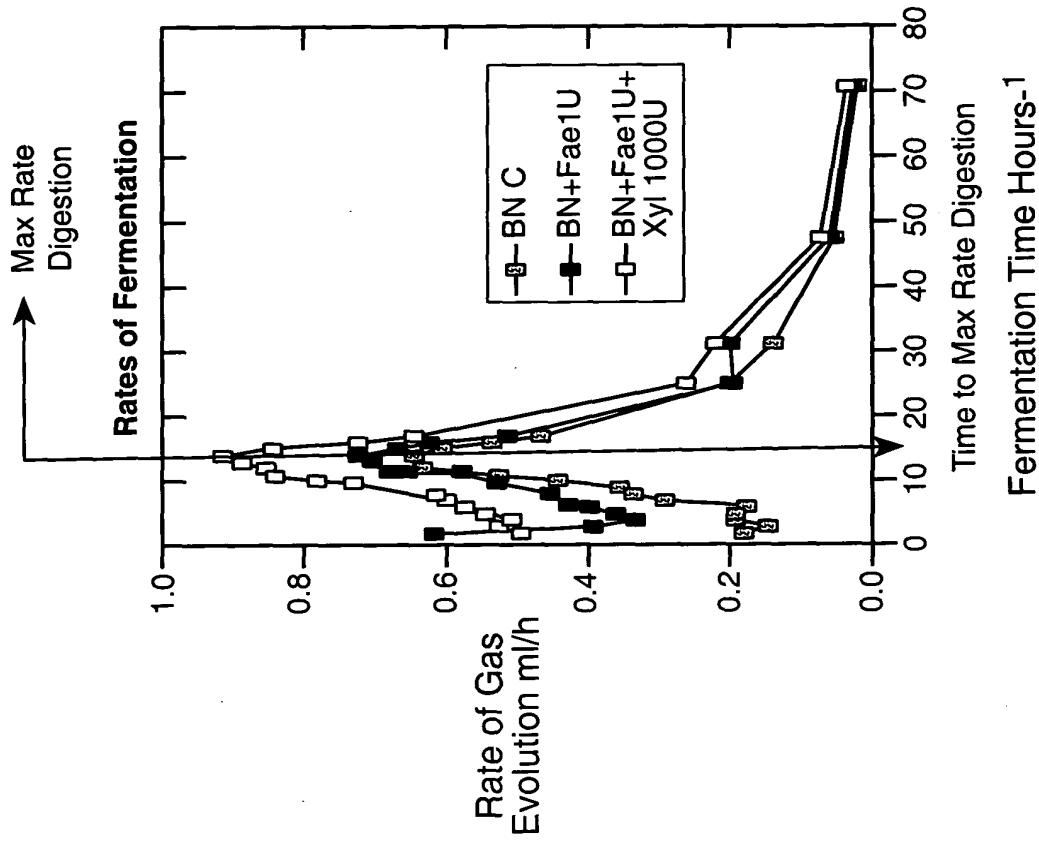


FIG.\_25B

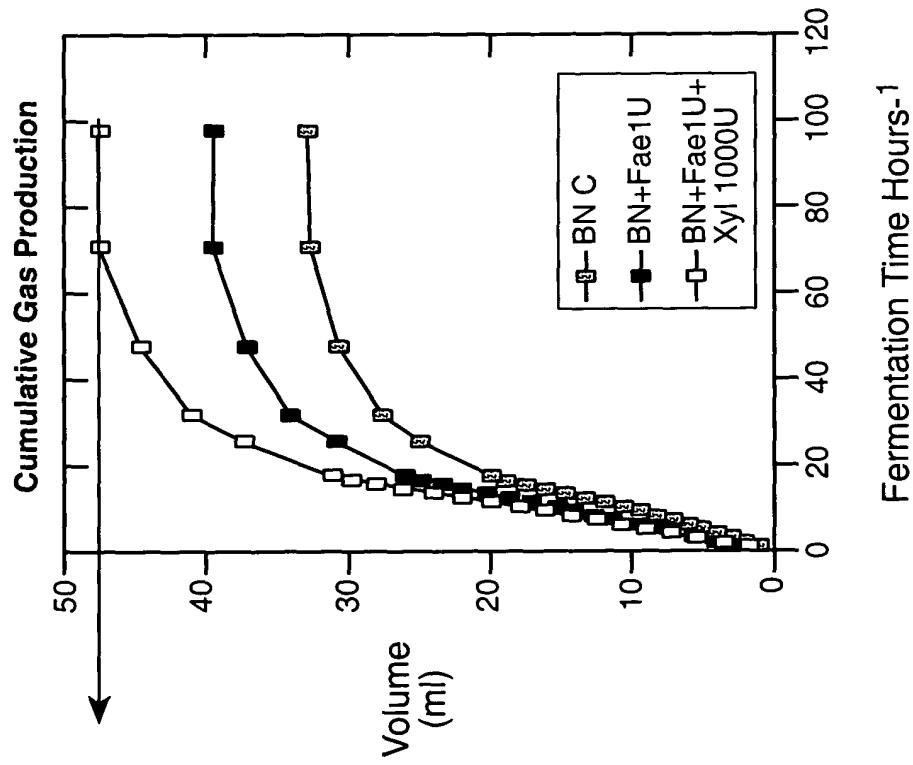
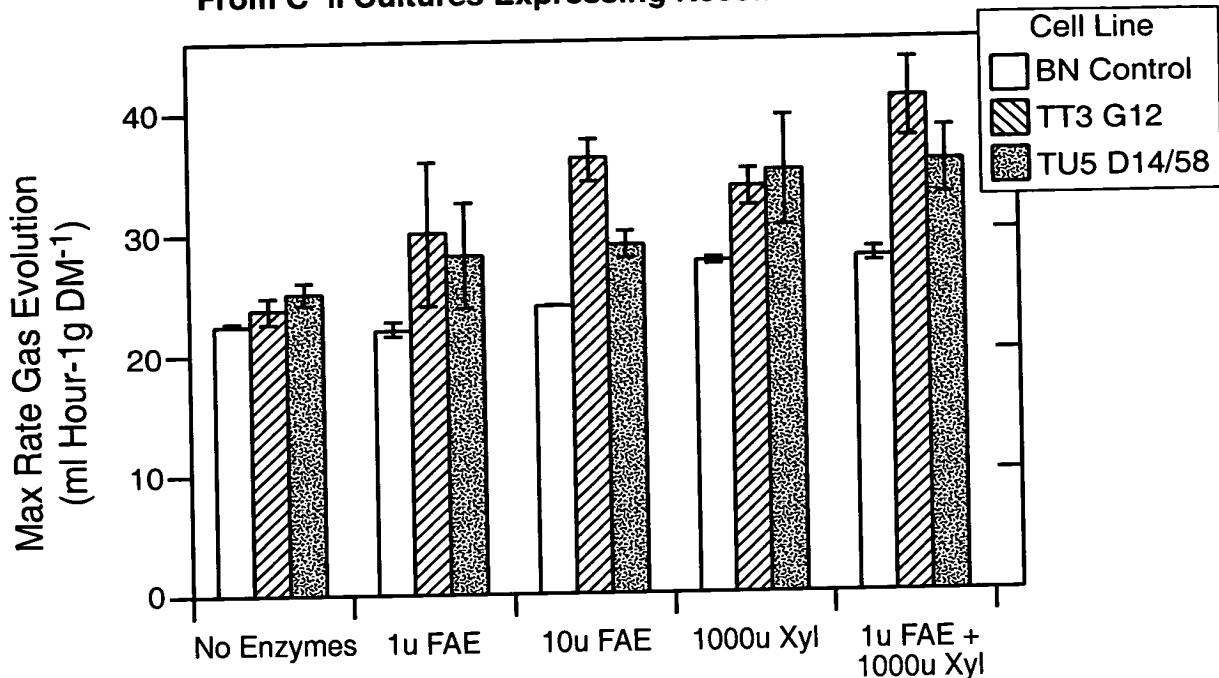


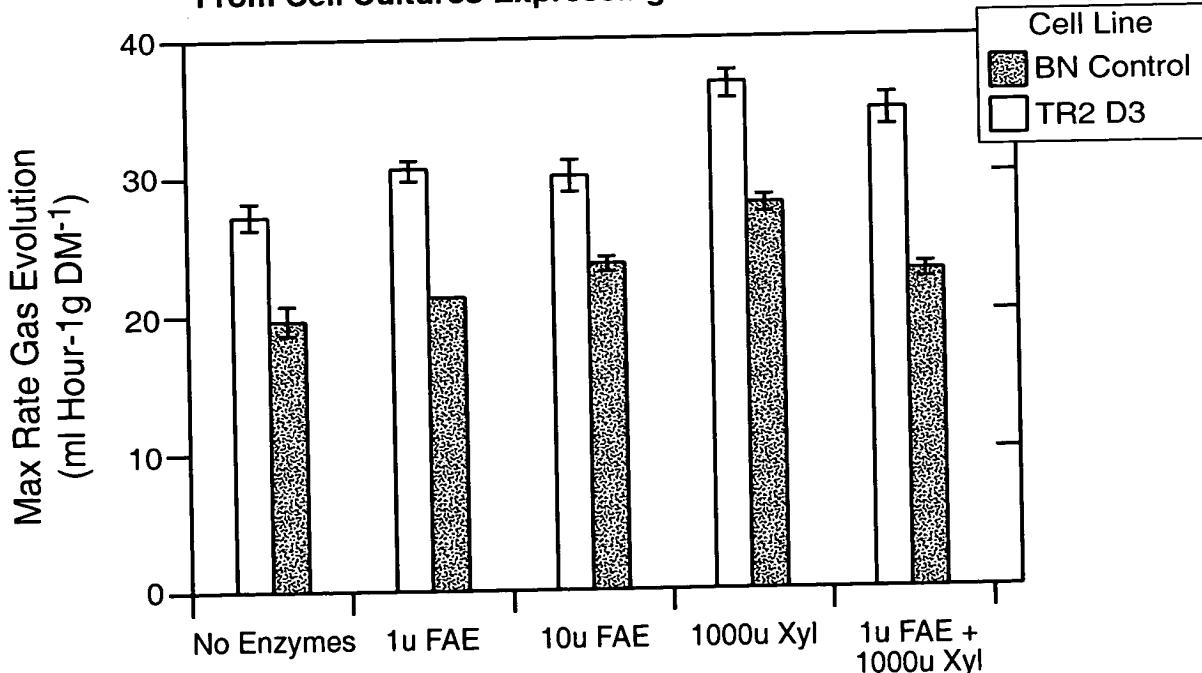
FIG.\_25A

**In-vitro Fermentation of *Festuca arundinacea* C II Walls  
From C II Cultures Expressing Recombinant FAE1**

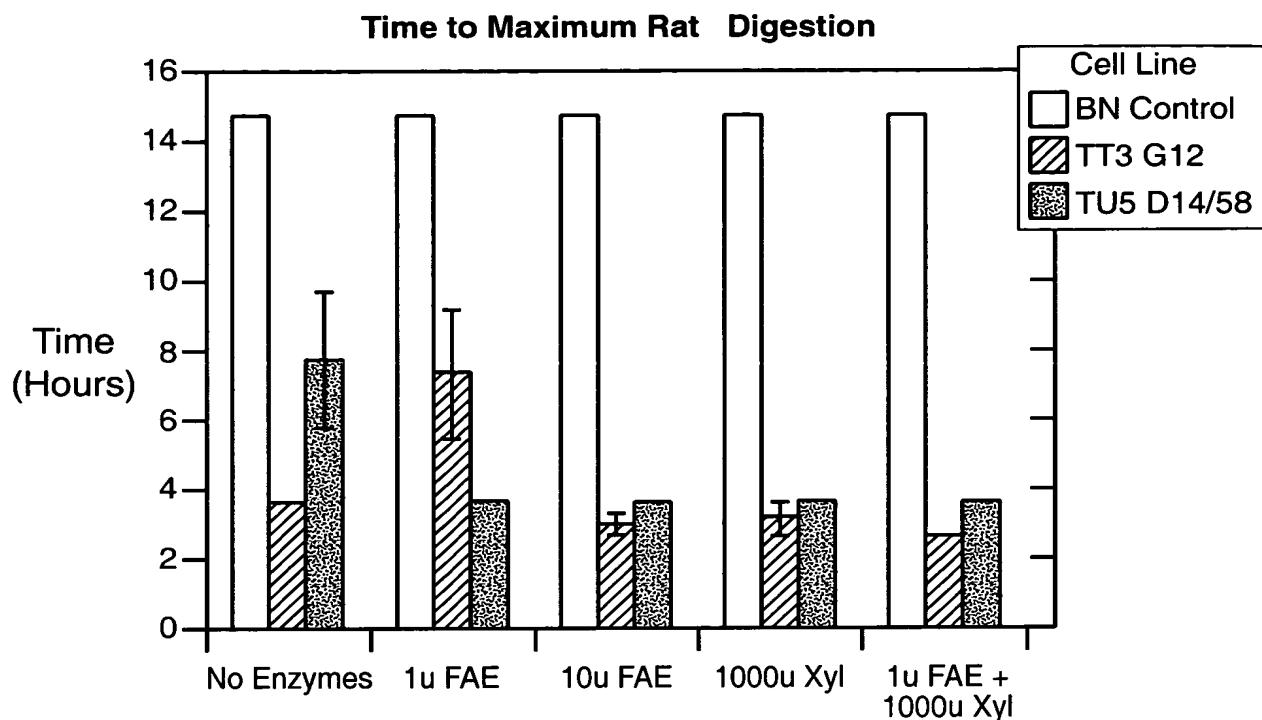


**FIG.\_26A** Maximum Rate of Digestion

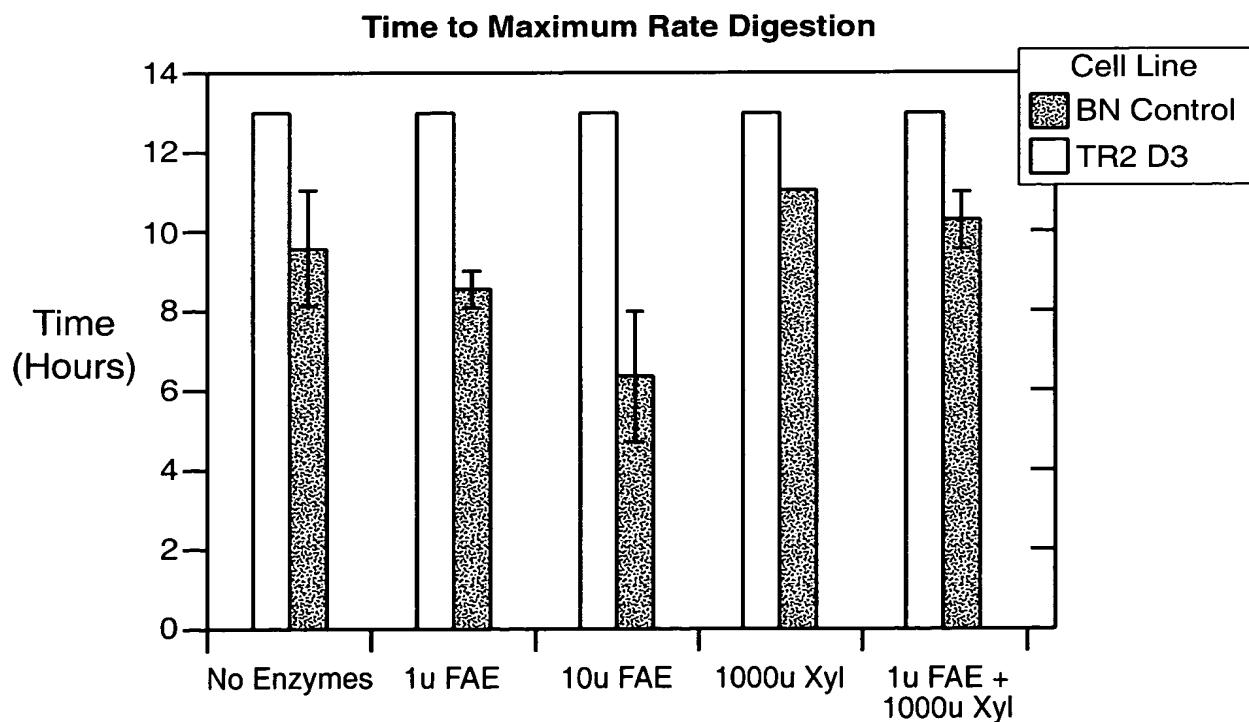
**In-vitro Fermentation of *Festuca arundinacea* Cell Walls  
From Cell Cultures Expressing Recombinant FAE1**



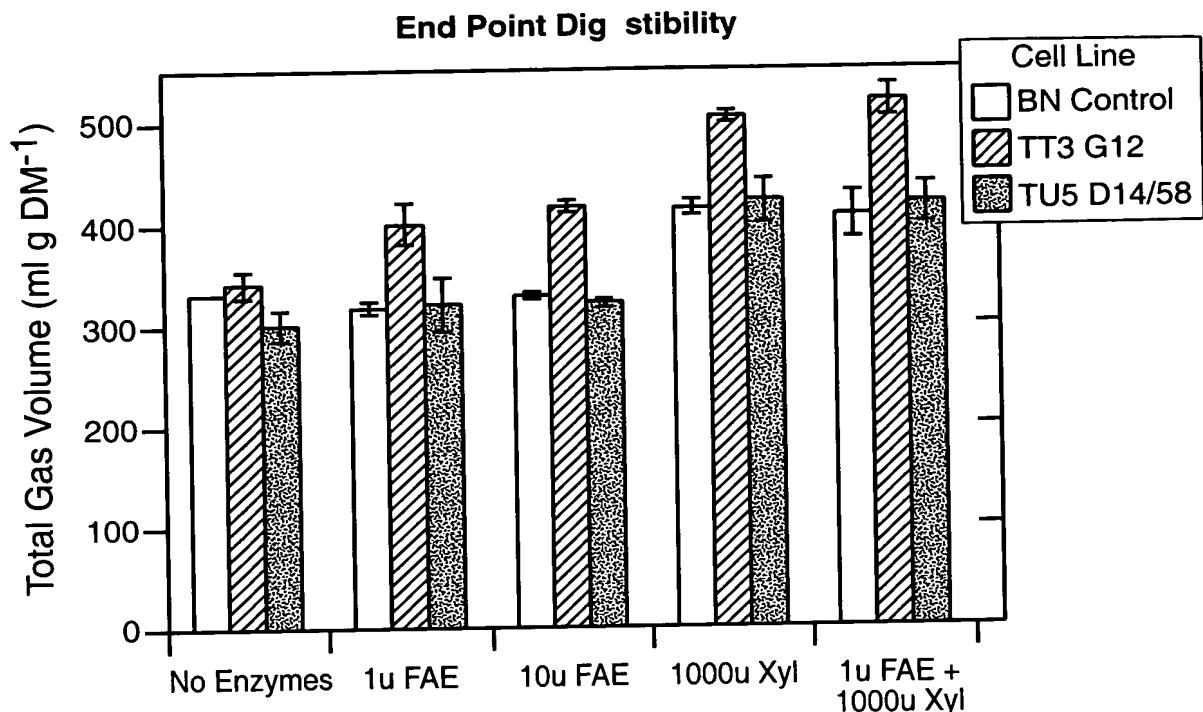
**FIG.\_26B** Maximum Rate of Digestion



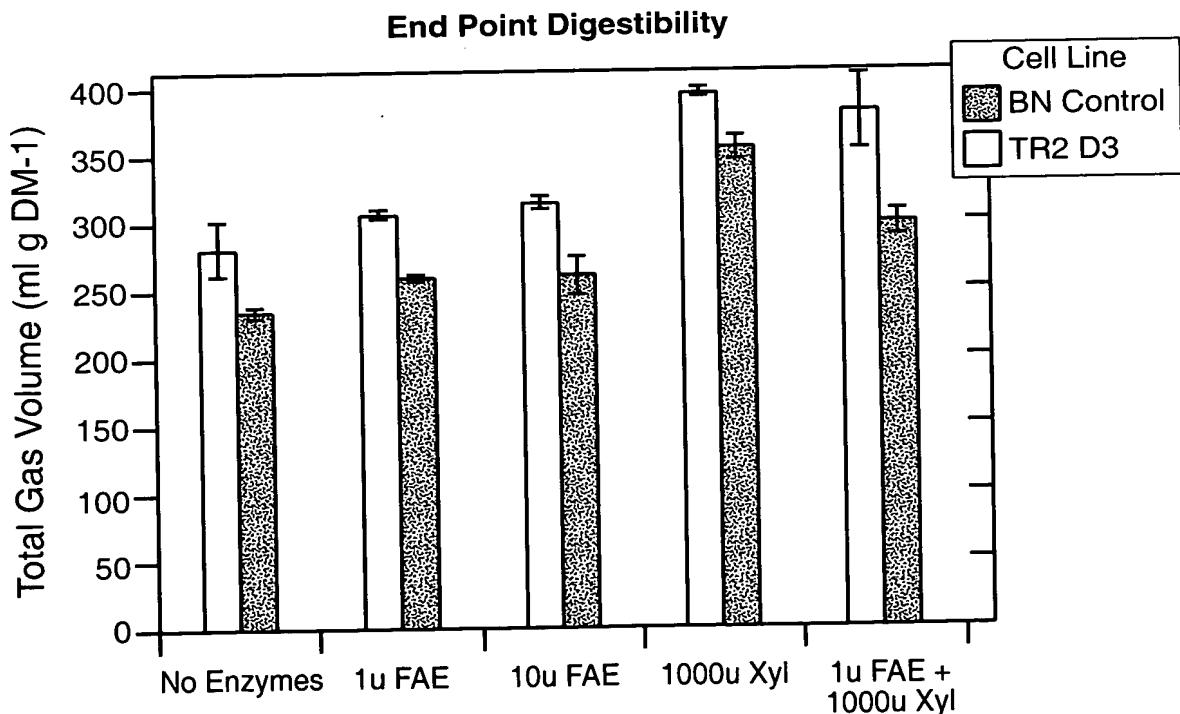
**FIG.\_27A**



**FIG.\_27B**

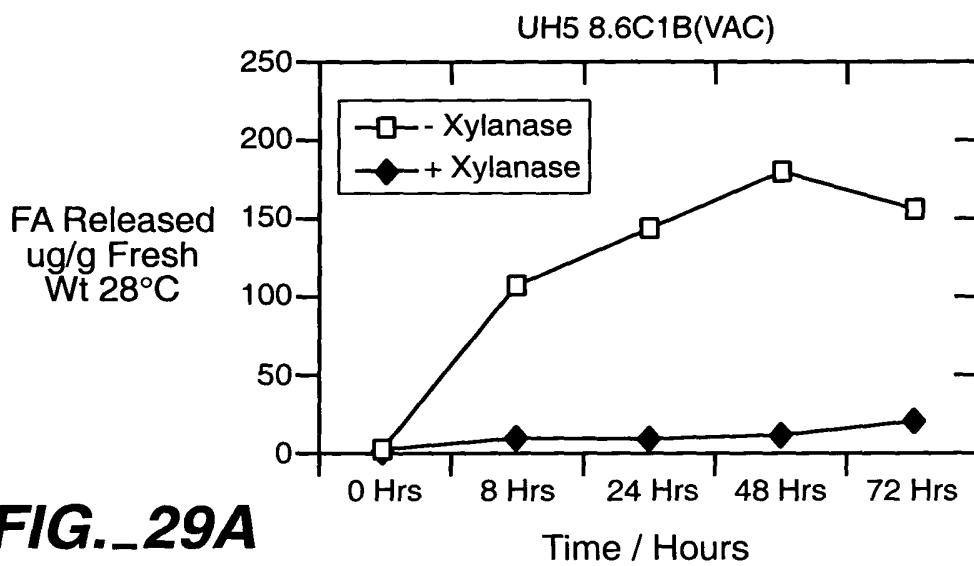


**FIG.\_28A**



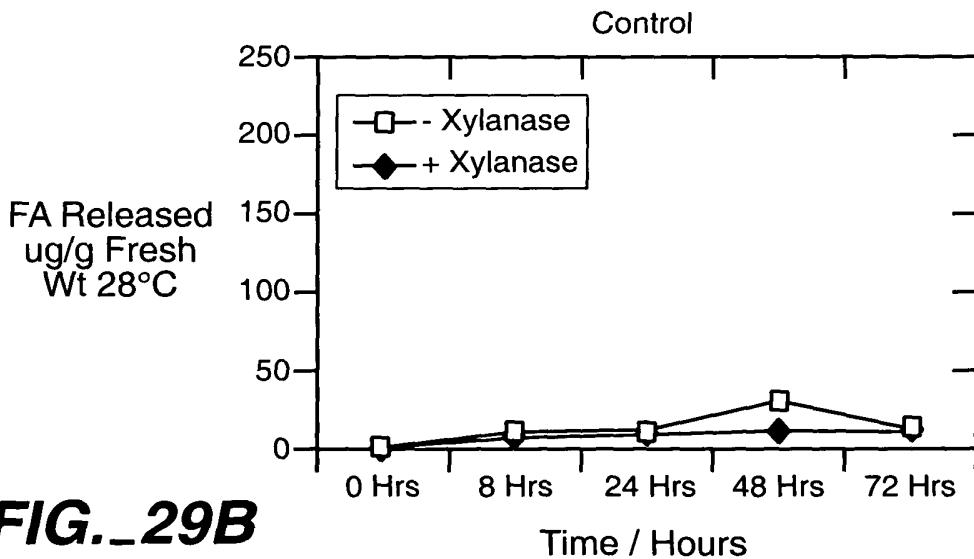
**FIG.\_28B**

**Kinetics of FAE Activity by Ferulic Acid Release  
from Cell Wall under Self Digestion in *Festuca arundinacea*  
and Stimulation by Xylanase**

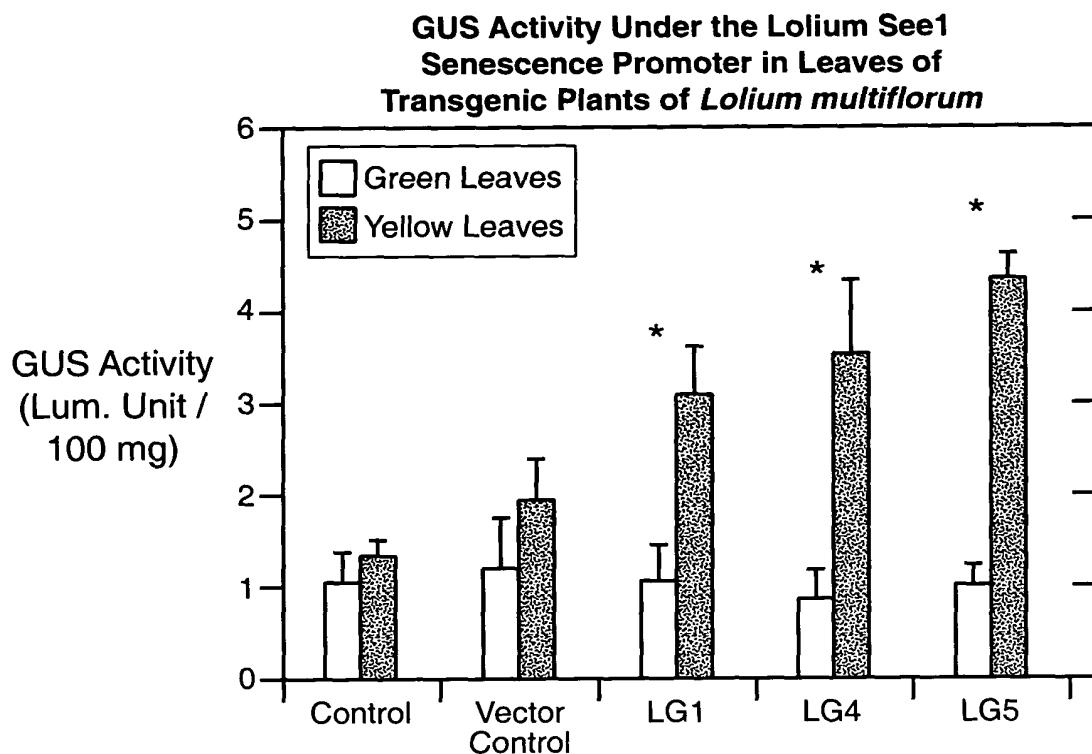


**FIG.\_29A**

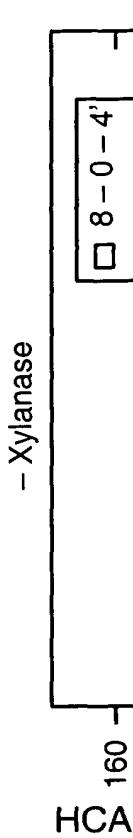
**Kinetics of FAE Activity by Ferulic Acid Release  
from Cell Wall under Self Digestion in *Festuca arundinacea*  
and Stimulation by Xylanase.**



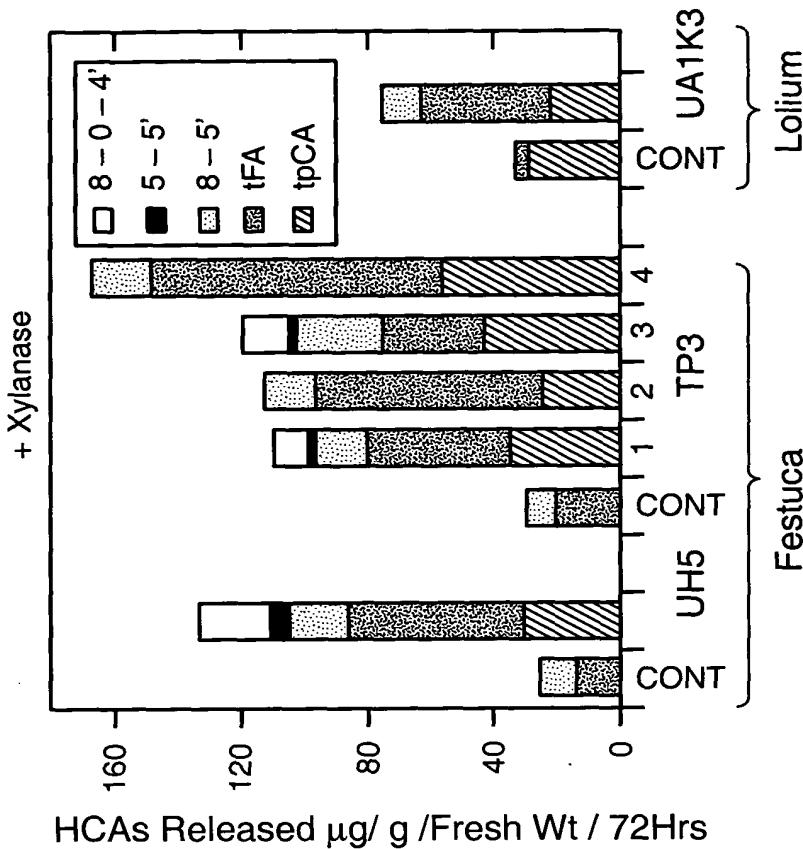
**FIG.\_29B**



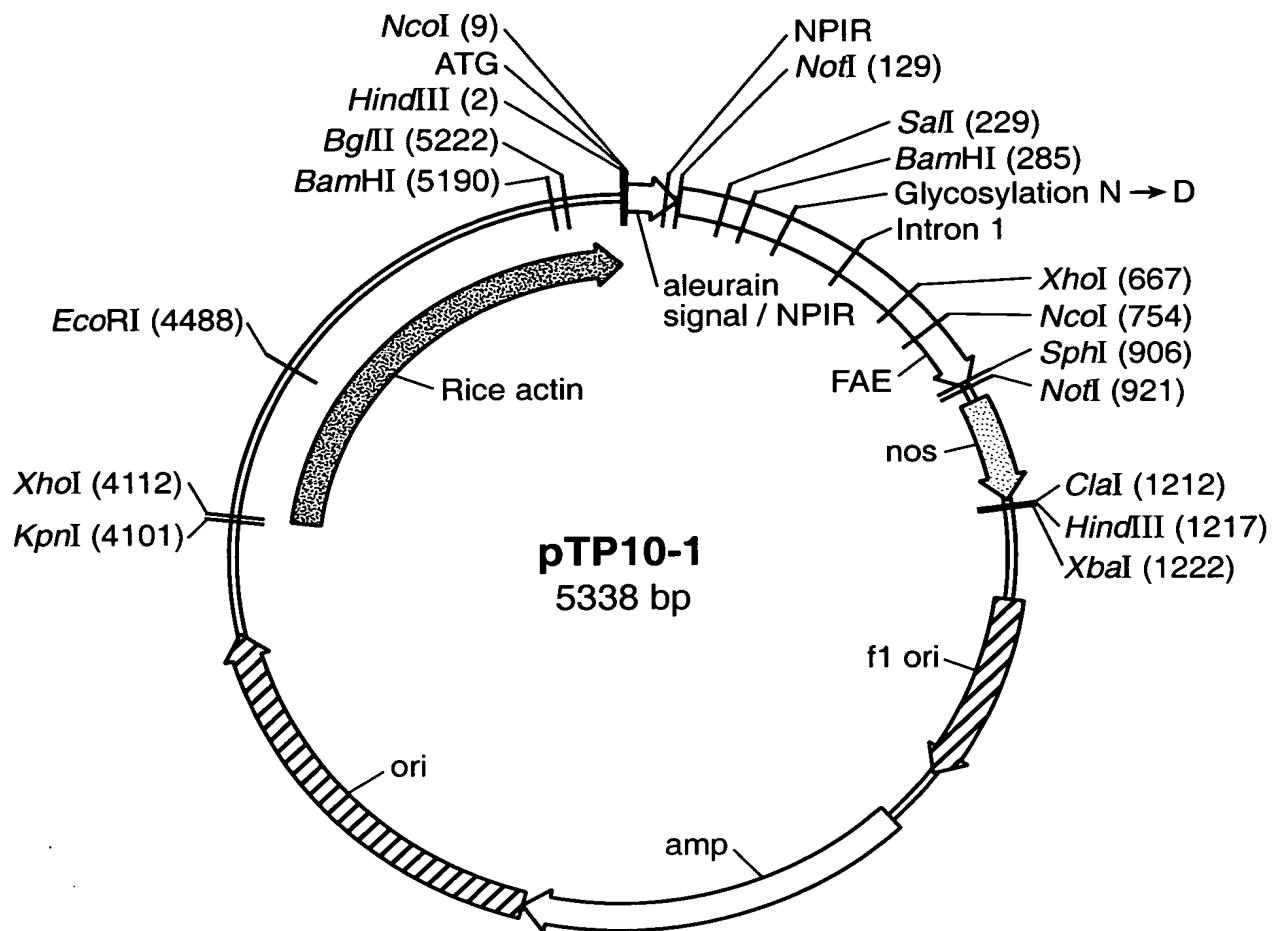
**FIG.\_30**



## FIG. 31A



## FIG. - 31B



**FIG.\_32A**

NcoI

HindIII

1 AAGCTTACCA TGGCCCAAGC CCGCGTCCCTC CTCCTGGCGC TCGCCGTGCT GCCCACGGCC GCGTCGCCG

NPIR

NotI

71 TCGCCTCCCTC CTCCCTCCCTC GCCGACTCCA ACCCGATCCG GCGCGTACCC GACCGCGCGG CGCGCTCCAC  
• Q G I S E D L Y S R L V E M A T I S Q A A Y A  
141 GCAGGGCATC TCCGAAGACC TCTACAGCCG TTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCTCTACGCC

SalI

211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTACAC CGTCTTCCGT GCCACTGGTA GTGATAACG  
BamHI

281 GATGGATCCT CGCGACGAC AGCAGCAAG ATTAAATCAC

Glycosylation

351 TCTACAACTC GATACTGACT ACACCCCTCAC GCCTTTCGAC ACCCTACCCAC AATGCAACGG TTGTGAAGTA  
H G G Y Y I G W V S V D Q V E S L V K Q Q V S  
421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTAA  
• Q Y P D Y A L T V T G H X L G A S L A A L T A  
491 GCCAGTATCC GGACTACGG CTGACCGTGA CGGGCCACKC CCTCGGGGCC TCCCTGGGG CACTCACTGCG  
• A Q L S A T Y D N I R L Y T F G E P R S G N Q  
561 CGCCCAAGCT TCTGCGACAT AGGACAACAT CGGACAACAT CGGCAATCAG

FIG.-32B

XbaI

A F A S Y M N D A F Q A S S P D T T Q Y F R V T  
631 GCCTTCGGGT CGTACATGAA CGATGCCCTTC CAAGCCTCGA GCCCAGATA GAGCAGTAT TTCCGGGTCA

H A N D G I P N L P P V E Q G Y A H G G V E Y  
701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCATGGCG GTGGTAGAGTA

W S V D P Y S A Q N T F V C T G D E V Q C C E  
771 CTGGAGCGTT GATCCTTACA GCGGCCAGAA CACATTGTC TGCACTGGG ATGAAAGTGCA GTGCTGTGAG

NcoI

A Q G G Q G V N N A H T T Y F G M T S G A C T W  
841 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGGGC GCATGCACCT

P V A A A E T T E G \*  
911 GGCGGTGCG GCGCGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAAA CATTTGGCAA

TAAAGTTCT TAAGATTGAA TCCCTGTTGCC GGTCTTGGA TGATTATCAT ATAATTTCGT TTGAATTACG  
981 TTAAGCATGT AATAATTAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTATGA TTAGAGTCCC  
1051 GCAATTATAC ATTAAATAC CGATAGAAA CAAATATAG CGCGAAACT AGGATAAATT ATCGCGCGCG

NotI

KDEL

TCTAGAAGCT TCGATAAGCT TCGTACTAGA TCGATAAGCT TCTAGAAGCT CCGGTGGAGG TCCAATTGCG CCTTATAGTGA  
981 CGTGTATTAC GCGCGCTAC TGGCGCTGT TTTACAACTGT CGTGACTGGG AAAACCTGG CGTACCCAA  
CTTAATCGCC TTGCGAGCAC A TCCCGCTTTC GCGAGCTGGC GTAATAGCGA AGAGGCCGC ACGATCGCC  
1051 CTTCCCAACA GTTGGCGAGC CTGAATGGG AATGGGACGC GCGCTGTAGC GGGCATTA GCGGGGGGG  
1121

HindIII

Clal

XbaI

1191 GTGTCACTA TGTTACTAGA TCGATAAGCT TCTAGAAGCT CCGGTGGAGG TCCAATTGCG CCTTATAGTGA  
1261 GTCGTATTAC GCGCGCTAC TGGCGCTGT TTTACAACTGT CGTGACTGGG AAAACCTGG CGTACCCAA  
1331 CTTAACCGCC TTGCGAGCAC A TCCCGCTTTC GCGAGCTGGC GTAATAGCGA AGAGGCCGC ACGATCGCC  
1401 CTTCCCAACA GTTGGCGAGC CTGAATGGG AATGGGACGC GCGCTGTAGC GGGCATTA GCGGGGGGG

FIG.-32C

Targeted 09/1209 - 07/02/02

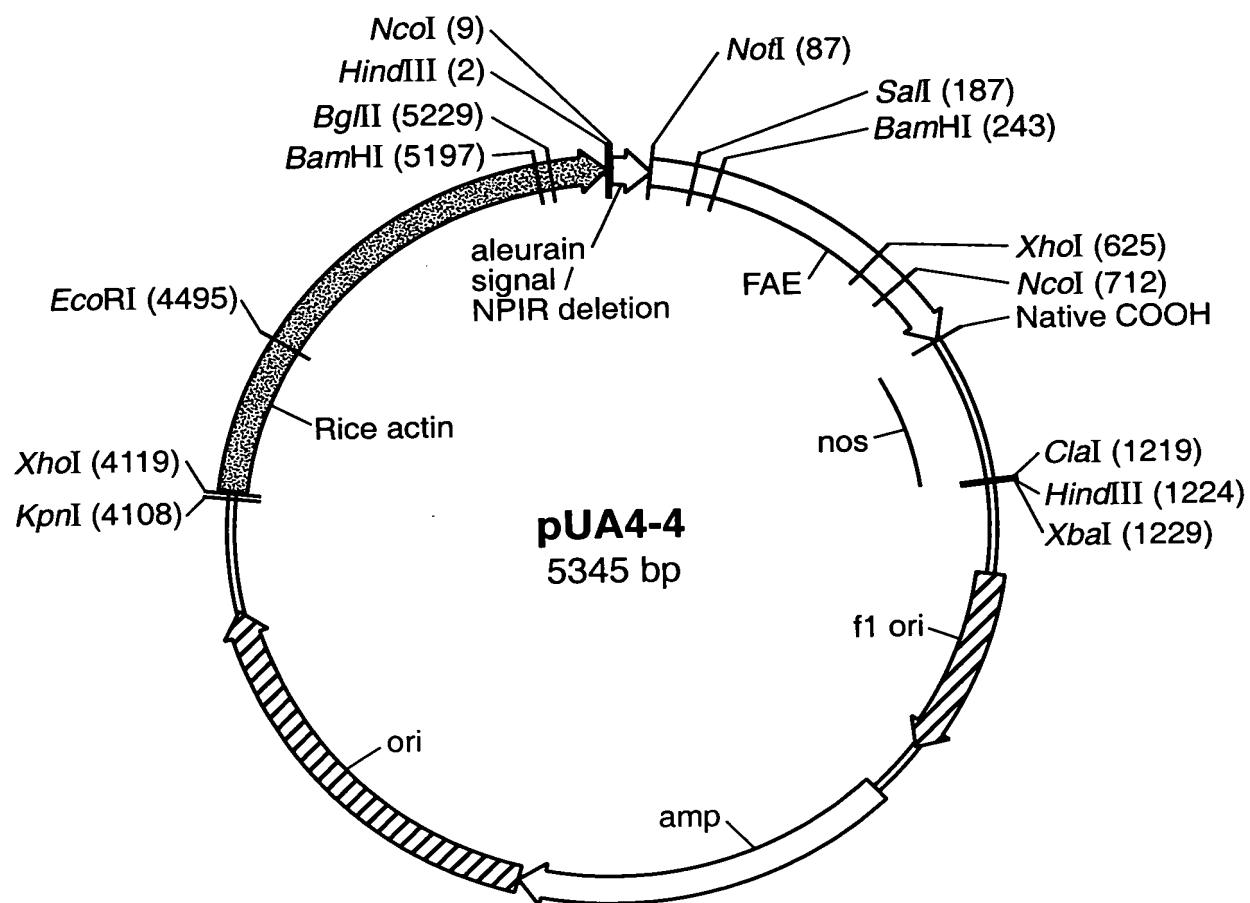
38 / 154

|      |              |             |             |             |             |             |             |
|------|--------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1471 | TGTGGTGGTT   | ACGGCGAGCG  | TGACCGCTAC  | ACTTGGCCAGC | GCCCTAGCCGC | CGCTTCCTTT  | CGCTTCCTT   |
| 1541 | CCTTCCTTTC   | TGCCAACGTT  | CGCCGGCTTT  | CCCCGTCAAG  | CTCTAACATCG | GGGGCTCCCT  | TTAGGGTTCC  |
| 1611 | GATTTAGTGC   | TTACGGCAC   | CTCGACCCCCA | AAAACCTGGA  | TTAGGGTGTAT | GGTCACGTA   | GTGGGCCATC  |
| 1681 | GCCCTGTGATAG | ACGGTTTTC   | GCCCCTTGAC  | GTTGGAGTCC  | ACGTTCTTTA  | ATAGTGGACT  | CTTGTGCCAA  |
| 1751 | ACTGGAAACAA  | CACTCAACCC  | TATCTGGTC   | TATTCTTTG   | ATTATAAGG   | GATTTGGCG   | ATTCTGGCCT  |
| 1821 | ATTGGTTAAA   | AAATGAGCTG  | ATTAAACAA   | AATTAAACGC  | GAATTAAAC   | AAAATATTAA  | CGCTTACAAT  |
| 1891 | TTAGGTGGCA   | CTTTTCGGGG  | AAATGTGGCG  | GGAACCCCTA  | TTTGTATT    | TTCCTAAATA  | CATTCAAATA  |
| 1961 | TGTATCCGCT   | CATGAGACAA  | TAACCCCTGAT | AAATGCTCA   | ATAATAATGGA | AAAAGGAAGA  | GTATGAGTAT  |
| 2031 | TCAACATTTTC  | CGTGTGCC    | TTATTCCTT   | TTTGGGCA    | TTTGCCTTC   | CTGTTTTGTC  | TCACCCAGAA  |
| 2101 | ACGCTGGTGA   | AAGTAAAGA   | TGCTGAAGAT  | CAGTGGGTG   | CACGAGTGGG  | TTACATCGAA  | CTGGATCTCA  |
| 2171 | ACAGCGGTAA   | GATCCTTGAG  | AGTTTTCGCC  | CCGAAGAACG  | TTTCCCAATG  | ATGAGCACTT  | TTAAAGTTCT  |
| 2241 | GCTATGTGGC   | GCGGTATTAT  | CCCGTATTGA  | CGCCGGCAA   | GAGCAACTCG  | GTGCGCGCAT  | ACACTATTC   |
| 2311 | CAGAATGACT   | TGGTTGAGTA  | CTCACCAAGTC | ACAGAAAAGC  | ATCTTACGGA  | TGGCATGACA  | GTAAGAGAAAT |
| 2381 | TATGCGAGTGC  | TGCCATAACC  | ATGAGTGTAA  | ACACTGGGC   | CAACTTACTT  | CTGACAACGA  | TGGAGGGACC  |
| 2451 | GAAGGGAGCTA  | ACCGCTTTT   | TGACAACAT   | GGGGATCAT   | GTAACTCGCC  | TTGATCGTGT  | GGAAACCGGAG |
| 2521 | CTGAATGAG    | COATACCAA   | CGACGAGCGT  | GACACACGA   | TGCCTGTAGC  | ATGGCAACAA  | ACGTTGGCAGA |
| 2591 | AACTATAAC    | TGGCGAACTA  | CTTACTCTAG  | CTTCCCGGCA  | ACAATTAAATA | GACTGGATGG  | AGGGGGATAAA |
| 2661 | AGTTCAGGGA   | CCACTTCTGC  | GCTGGCCCT   | TCCGGCTGGC  | TGTTTATTG   | CTGATAAAATC | TGGAGCCGGCT |
| 2731 | GAGCGTGGGT   | CTCGGGGTAT  | CATGGCAGCA  | CTGGGGCCAG  | ATGTTAAGCC  | CTCCCGTATC  | GTAGTTATCT  |
| 2801 | ACACGACGGG   | GAGTCAGGCA  | ACTATGGATG  | AACGAAATAG  | ACAGATCGCT  | GAGATAGGTG  | CCTCACTGAT  |
| 2871 | TAAGGATTTGG  | TAACGTGTAG  | ACCAAGTTTA  | CTCATATATA  | CTTTAGATTG  | ATTAAAACCT  | TCATTTTTAA  |
| 2941 | TTTAAAGGA    | TCTAGGTGAA  | GATCCTTTT   | GATAATCTCA  | TGACCAAAAT  | CCCTTAACGT  | GAGTTTCGTT  |
| 3011 | TCCACTGAGC   | GTCAAGACCC  | GTAGAAAAGA  | TCAAAGGATC  | TTCTTGAGAT  | CCTTTTTTTC  | TGCGCGTAAAT |
| 3081 | CTGCTGTTG    | CAAACAAAAA  | AACACCGCT   | ACCGGGTG    | GTTTGTTGTC  | GGGATCAAGA  | GCTACCAACT  |
| 3151 | CTTTTCGGA    | AGGTAACCTG  | CTTCAGGAGA  | GCGCAGATAAC | CAAATACTGT  | CCCTCTAGTG  | TAGCCGTAGT  |
| 3221 | TAGGCCACCA   | CTTCAGAAC   | TCTGTAGCAC  | CCCTACATA   | CCTCGCTCTG  | CTAATCCTGT  | TACCAAGTGGC |
| 3291 | TGCTGCCAGT   | GGCGATAAGT  | CGTGTCTTAC  | CGGGTTGGAC  | TCAAGACGAT  | AGTTACCGGA  | TAAGGGCGAG  |
| 3361 | CGGTGGGCT    | GAACGGGGGG  | TTCGTGCACA  | CAGCCAGCT   | TGGAGCGAAC  | GACCTACACC  | GAACCTGAGAT |
| 3431 | ACCTACAGCG   | TGAGCTATGA  | GAAAGGCCA   | CGCTTCCGA   | AGGGAGAAAG  | GCGGACAGGT  | ATCCGGTAAG  |
| 3501 | CGGCAGGGTC   | GGAACAGGGAG | AGGCCACGG   | GGAGCTTCCA  | GGGGAAACAG  | CCTGGTATCT  | TTATAGTCCT  |
| 3571 | GTGGGGTTTC   | GCCACCTCTG  | ACTTGAGCGT  | CGATTTTGT   | GATGCTCGTC  | AGGGGGGG    | AGCCTATGGA  |
| 3641 | AAAACGCCAG   | CAACGCCAG   | TCTTACGGT   | TCCTGGCCCT  | TTGCTGGCT   | TTTGCTCA    | TGTCTTTTC   |
| 3711 | TGGCTTATTC   | CCTGATTCTG  | TGGATAACCG  | TATTAACGCC  | TTTGAGTGAG  | CTGATAACGC  | TGGCCGCAGC  |
| 3781 | CGAACGACCC   | AGGCCAGCGA  | GTCAACTGAGC | GAGGAAGCGG  | AAGAGGCC    | AATACGCAA   | CCGGCTCTCC  |
| 3851 | CCGGCGGTTG   | GCCGATTCA   | TAATGCGACT  | GGCACGACAG  | GTTCCTCCGAC | TGGAAAGCGG  | GCAGTGAGCG  |

FIG.-32D

|      |             |             |             |             |             |             |              |       |
|------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|-------|
| 3921 | CAACGCAATT  | AATGTGAGTT  | AGCTCACTCA  | TTAGGCACCC  | CAGGCTTAC   | ACTTTATGCT  | TCCGGCTCGT   |       |
| 3991 | ATGTTGTGTG  | GAATTGTGAG  | CGGATAACAA  | TTTCACACAG  | GAACAGCTA   | TGACCATGAT  | TACGCCAAGC   |       |
|      |             |             |             |             |             |             |              | KpnI  |
|      |             |             |             |             |             |             |              | XbaI  |
| 4061 | GCGCAATTAA  | CCCTCACTAA  | AGGGAAACAA  | AGCTGGGTAC  | CGGGCCCC    | CTCGAGGTCA  | TTCATATGCT   |       |
| 4131 | TGAGAAGAGA  | GTGGGGATAG  | TCCAAAATAA  | AACAAAGGTAA | AGATTACCTG  | GTCAAAAGTG  | AAAACATCAG   |       |
| 4201 | TTAAAAGGTG  | GTATAAGTAA  | AATATCGGTAA | ATAAAAGGTG  | GCCCCAAAGTG | AAATTTACTC  | TTTCTACTA    |       |
| 4271 | TTATAAAAAT  | TGAGGATGTT  | TTGTCGGTAC  | TTTGATAGTGT | CATTTTGTA   | TGAATTGGTT  | TTTAAGTTA    |       |
| 4341 | TTCGGGAAATT | GGAAATGGCAT | ATCTGTATT   | GAGTGGTTT   | TTAAGTTCGT  | TGCTTTTGTAA | AAATACAGAGG  |       |
| 4411 | GATTGTATA   | AGAAATATCT  | TTAAAAAACC  | CATATGCTAA  | TTTGACATAA  | TTTTTGAGAA  | AAATATATA    |       |
|      |             |             |             |             |             |             |              | ECORI |
| 4481 | TCAGGGCAAT  | TCCACAAATGA | ACAAATAATA  | GATTAAATA   | GCTTGCCCC   | GTGCGAGCGA  | TGGTATT      |       |
| 4551 | TTCTAGTAAA  | ATAAAAGATA  | AACTTAGACT  | CAAAACATT   | ACAAAAACAA  | CCCCTAAAGT  | CCTAAAGCCC   |       |
| 4621 | AAAGTGCAT   | GCACGATCCA  | TAGCAAGCCC  | AGCCCCAACCC | AACCCCAACCC | AGTGCAGCCA  |              |       |
| 4691 | ACTGGCAAT   | AGTCTCCACC  | CCGGGCACTA  | TCACCGTGTAG | TTGTCGGCAC  | CACCGCACGT  | CTCGCAGCCA   |       |
| 4761 | AAAAAAAAGAA | AAAAGAAA    | AAAAGAAA    | AGAAAACAG   | CAGGTGGTCA  | CGGGTCGTGG  | GGGGCGGAA    |       |
| 4831 | AGCGAGGAGG  | ATCGCGAGCA  | GCGACGAGGC  | CGGGCCCTCC  | CTCCGTTCC   | AAAGAAACGC  | CCCCCATCGC   |       |
| 4901 | CACTATAC    | ATACCCCCC   | CTCTCCTCCC  | ATCCCCCAA   | CCCTACCAAC  | ACCACCAACCA | CCACCTCCCT   |       |
| 4971 | CCCCCTCGCT  | GCGGACGAC   | GAGCTCCCTCC | CCCTCCCCC   | TCCGCCGCC   | CCGGTAACCA  | CCCGCCCT     |       |
| 5041 | CTCCCTCTTC  | TTTCTCCGTT  | TTTTTTTTCG  | TCTCGGTCTC  | GATCTTTGGC  | CTTGGTAGTT  | TGGGTGGGCG   |       |
| 5111 | AGAGCGGCTT  | CGTCGCCAG   | ATCGGTGCGC  | GGGAGGGCG   | GGATCTCGCG  | GCTGGCGTCT  | CCGGGGGTGA   |       |
|      |             |             |             |             |             |             |              | BglII |
| 5181 | GTGGCCCGG   | ATCCCTCGGG  | GGATGGGGC   | TCTCGGATGT  | AGATCTCTT   | TCTTTCTCTT  | TTTGTGGTA    |       |
| 5251 | GAATTGAAAT  | CCCTCAGCAT  | TGTTCATCGG  | TAGTTTTCT   | TTTCATGATT  | TGTGACAAT   | GCAGGCCCTCGT |       |
| 5321 | GGGGAGCTTT  | TTTGTAGAC   |             |             |             |             |              |       |

FIG.-32E



**FIG.\_33A**

2009.07.02 0202

41 / 154

NcoI

HindIII

M A H A R V L L L A L A V L A T A A V A V  
 1 AAGCTTACCA TGGCCCAACGC CGCGCTCCTC CTCCCTGGCGC TCGCCGTGCT GGCCACGGCC GCGTCGCCG

NcoI

• A S S R A A S T Q G I S E D L Y S R L V E M .  
 71 TCGCCTCCTC CGCGCGGCC GCCTCCACGC AGGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAATAT  
 Sall

• A T I S Q A A Y A D L C N I P S T I I K G E K .  
 141 GGCCACTATC TCCAAAGCTG CCTACGGCGA CCTGTGCAAC ATTCCGTGCA CTATTATCAA GGGAGAGAAA  
 BamHI

I Y N S Q T D I N G W I L R D D S S K E I I T V  
 211 ATTACAAATT CTCAAACCTGA CATTAAACGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA ATAATCACCG  
 • F R G T G S D T N L Q L D T N Y T L T P F D T .  
 281 TCTTCGGGG CACTGGTAGT GATACTGAACTC TACAACCTCGA TACTAACTAC ACCCTCTCACGC CTTTCGACAC  
 • L P Q C N G C E V H G Y I G W V S V Q D Q .  
 351 CCTACCAAA TGCAACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA  
 V E S L V K Q Q V S Q Y P D Y A L T V T G H X L  
 421 GTCGAGTGC GCAGGTTAGC CAGTATCGG ACTACGGCT GACCGTGAAC GGCCACKCCC  
 • G A S L A A L T A A Q L S A T Y D N I R L Y T .  
 491 TCGGGCCTC CCTCACTGGCGA CTCAGCTGTC CCCAGCTGTC TGGACATAC GACAACATCC GCCTGTACAC  
 XbaI

• F G E P R S G N Q A F A S Y M N D A F Q A S S .  
 561 CTTGGCGAA CGGGCAGCG GCAATCAGGC CTTGGTCTG TACATGAAAG ATGCCTTCGA AGCCTCGAGC  
 P D T T Q Y F R V T H A N D G I P N L P P V E Q .  
 631 CCAGATACGA CGCAGTATT CGGGTCACT CATGCCAACG ACGGCAATCCC AACCTGGCC CGGTGGAGC  
 NcoI

• G Y A H G G V E Y W S V D P Y S A Q N T F V C .

FIG.-33B

|     |               |             |               |             |              |             |             |
|-----|---------------|-------------|---------------|-------------|--------------|-------------|-------------|
| 701 | AGGGGTACGC    | CCATGGCGGT  | GTAGAGTACT    | GGAGCGGTGA  | TCCTTACAGC   | GCCCCAAACA  | CATTGGTCTG  |
| .   | T G D E V Q C | C E A Q G G | Q G V N N A H | T T Y       |              |             |             |
| 771 | CACTGGGAT     | GAAGTGCAGT  | GCTGTAGGGC    | CCAGGGGGGA  | CAGGGGTGGA   | ATAATGGGCCA | CACGACTTAT  |
|     | F G M T S G A | C T W       | *             |             |              |             |             |
| 841 | TTTGGGATGA    | CGAGGGGAGC  | CTGTACATGG    | TGATCAGTCA  | TTTCAGGCCCTC | CCCGAGGTGA  | CCAGGAAGAA  |
| 911 | TGGATGTCCT    | GGAGGGGGG   | CCGGTAAACC    | ACTGAAAGGAT | GAGCTGTAAA   | GAAGCAGATC  | GTTCAAAACAT |
| 981 | TRGCAATAA     | AGTTCTAA    | GATTGAATCC    | CTTGGCGGT   | CTTGCATGAA   | TTATCATATA  | ATTTCGTGTT  |
| 051 | AATTAGTTA     | AGCATGTAAT  | AATTAAACATG   | TAATGCATGA  | CGTTATTAT    | GAGATGGGT   | TTTATGATTAA |
| 121 | GAGTCCCGCA    | ATTATACATT  | TAATACCGA     | TAGAAAACAA  | AATATAGGCC   | GCAAACTAGG  | ATAAATTATC  |

|       | XbaI        | Clai        |            |
|-------|-------------|-------------|------------|
| 11191 | GCGCGGGGTG  | TCATCTATGT  | TACTAGATCG |
| 11261 | ATAGTGAAGTC | GTATTACGG   | ATAAGCTTCT |
| 11331 | TACCCAACTT  | CGCTCACTGG  | AGAGGGCCG  |
| 11401 | GATGCCCTT   | CCGACATCC   | ACAAAGTCGT |
| 11471 | CGGGGGGT    | CCCAACAGTT  | CCCTTCGCC  |
| 11541 | TTCTTCCCT   | GCGAGGCTG   | AGCTGGCTA  |
| 11611 | GGGTTCCGAT  | GGTGGTTACG  | GGGACGGCC  |
| 11681 | GGCCATGCC   | CCACGTTCTG  | GGGACGGCC  |
| 11751 | GTTCATGCC   | CGCAGGTGA   | GGGACGGCC  |
| 11821 | TCGGCCTATT  | GGAAACAAC   | GGGACGGCC  |
| 11891 | TTACAATT    | TCAACCCCT   | GGGACGGCC  |
| 11961 | TCAAATATGT  | GGTGGCACTT  | GGGACGGCC  |
| 2031  | TGAGTATTCA  | ATCCGGCTCAT | GGGACGGCC  |
| 2101  | CCAGAAACG   | ACATTTCGGT  | GGGACGGCC  |
| 2171  | GATCTAACCA  | CTGGTAAAGAT | GGGACGGCC  |
| 2241  | AAGTTCTGCT  | CCTTGAGAGT  | GGGACGGCC  |
| 2311  | CTATTCTCAG  | GGGTTAAAGAT | GGGACGGCC  |
| 2381  | AGAGAATTAT  | GTATTATCCC  | GGGACGGCC  |
| 2451  | GAGGACCCAA  | ATGAGCTTGC  | GGGACGGCC  |
| 2521  | ACCGGAGCTG  | TTGAGTACTC  | GGGACGGCC  |
| 2591  | TGGCCAAAC   | CATAACCATG  | GGGACGGCC  |
| 2661  | CGATAAAGT   | AGTGAATGCC  | GGGACGGCC  |

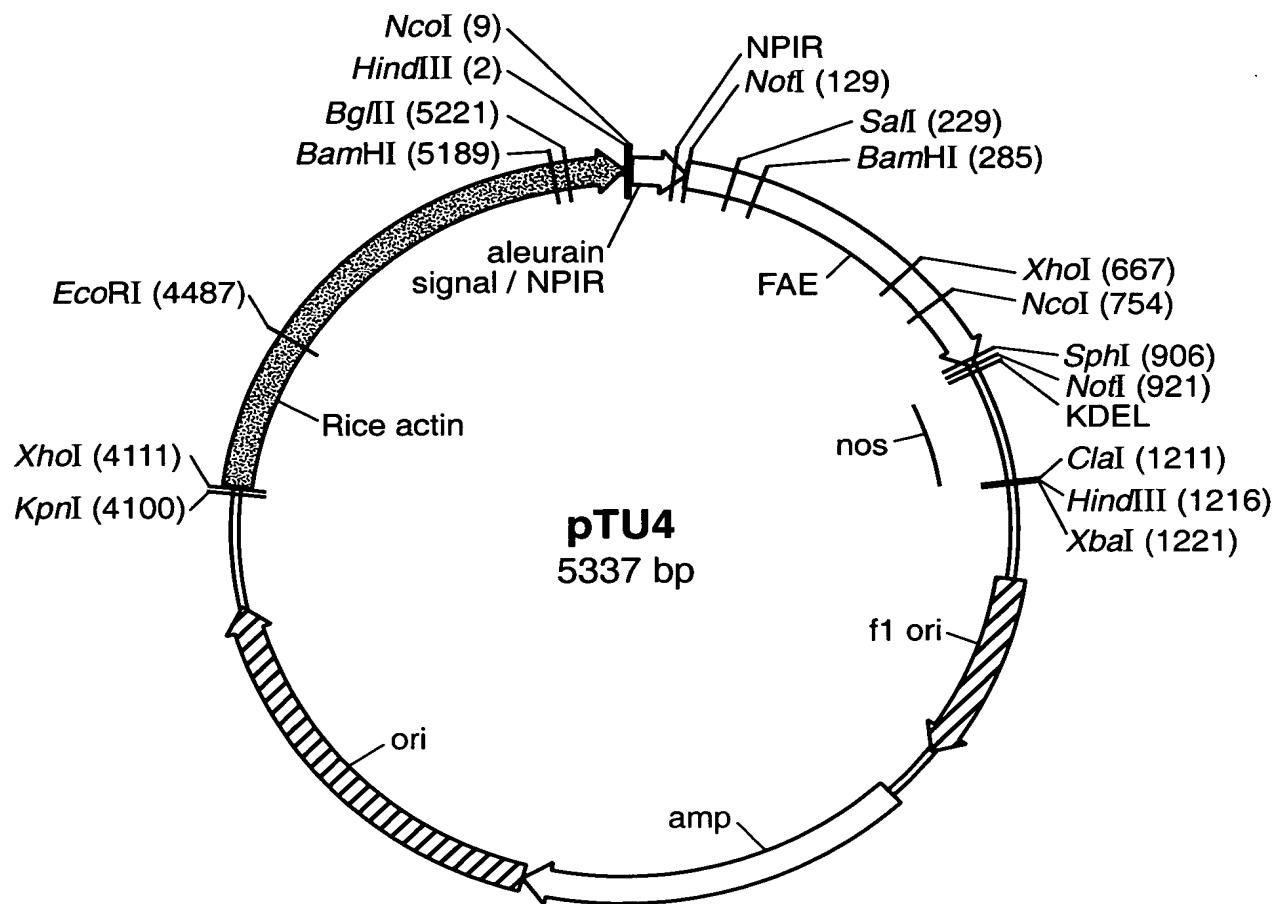
FIG. 33C

|      |             |             |             |             |             |             |             |
|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 2731 | AGCGGGTGTAG | CGGGTATCTC  | GCGGTATCAT  | TGGCAGCACTG | GGGCCAGATG  | GTAAGCCCTC  | CCGGTATCGTA |
| 2801 | GTTATCTACA  | CGACGGGGAG  | TCAGGCAACT  | ATGGATGAAAC | GAAATAGACA  | GATCGCTGAG  | ATAGGTGCCT  |
| 2871 | CACTGTTAA   | GCATGGTAA   | CTGTCAGACC  | AAAGTTACTC  | ATATATACTT  | TAGATTGATT  | TAAAACTTCA  |
| 2941 | TTTTAATT    | AAAAGGATCT  | AGGTGAAGAT  | CCTTTTGAT   | AATCTCATGA  | CCAAATCCC   | TTAACGTGAG  |
| 3011 | TTTCGTTCC   | ACTGAGGGTC  | AGACCCCGTA  | AAAAAGATCA  | AAGGATCTTC  | TGAGATCTCT  | TTTTTCTGC   |
| 3081 | GCGTAATCTG  | CTGCTTGAA   | ACAAAAAAAC  | CACCGCTACC  | AGCGGTGGTT  | TGTTTGCCTG  | ATCAAGAGGCT |
| 3151 | ACCAACTCTT  | TTTCCGAGG   | TAACTGGCTT  | CAGCGAGGG   | CAGATACCAA  | ATACTGTCTT  | TCTAGTGTAG  |
| 3221 | CCGTAGTTAG  | GCCACCCTT   | CAAGAACTCT  | GTAGCACCGC  | CTACATACCT  | CGCTCTGCTA  | ATCCTGTTAC  |
| 3291 | CAGTGGCTGC  | TGCCAGTGGC  | GATAAGTCGT  | GTCTTACCGG  | GTGGGACTCA  | AGACGATACT  | TACCGGATAA  |
| 3361 | GGGCCAGCGG  | TGGGCTGAA   | CGGGGGTTTC  | GTGCACACAG  | CCCAGCTTGG  | AGCGAACGAC  | CTACACCGAA  |
| 3431 | CTGAGATACC  | TACAGGTGA   | GCTATGAGAA  | AGCGGCCACGC | TTCCCGAAGG  | GAGAAAGGGG  | GACAGGTATC  |
| 3501 | CGGTAAGCGG  | CAGGGTGGA   | ACAGGAGAGC  | GCACGGAGGA  | GCTTCCAGGG  | GAAAACGCC   | GGTATCTTTA  |
| 3571 | TAGTCCGTGTC | GGGTTTCGCC  | ACCTCTGACT  | TGAGCGTCGA  | TTTTTGTGAT  | GCTCGTCAGG  | GGGGGGAGGC  |
| 3641 | CTATGGAAA   | ACGCCAGCAA  | CGGGCCCTTT  | TTACCGTTCC  | TGGCCCTTTG  | CTGGCCCTTTT | GCTCACATGT  |
| 3711 | TCTTTCTGCG  | GTATTCCTCT  | GATTCGTGG   | ATAACCGTAT  | TACCGCTTT   | GAGTGAGCTG  | ATACCGCTCG  |
| 3781 | CCGCAGCCGA  | ACGACCGAGC  | GCAGCGAGTC  | AGTGAGCGAG  | GAAGGGGAAG  | AGCGCCCAAT  | ACGCAAACCG  |
| 3851 | CCTCTCCCCG  | CGCGTTGGCC  | GATTCATTAA  | TGCAAGCTGGC | ACGACAGGTT  | TCCCAGCTGG  | AAAGCGGGCA  |
| 3921 | GTGAGGCCAA  | CGCAATTAAAT | GTGAGTTAGC  | TCACTCATTA  | GGCACCCAG   | GCTTACACT   | TTATGCTTC   |
| 3991 | GGCTCGTATG  | TGTGTGGAA   | TTGTGTGGCG  | ATAACAATT   | CACACGGAA   | ACAGCTATGA  | CCATGATTAC  |
|      |             |             |             |             | KpnI        | ~~~~~       |             |
|      |             |             |             |             | XbaI        | ~~~~~       |             |
| 4061 | GCCAAGGGCG  | CAATTAACCC  | TCACTAAAGG  | GAACAAAGC   | TGGGTACCGG  | GGCCCCCTC   | GAGGTCAATT  |
| 4131 | ATATGCTTGA  | GAAGGAGGTC  | GGGATAGTCC  | AAATAAAAC   | AAAGGTAAAGA | TACCTGGTC   | AAAAGTGAAGA |
| 4201 | ACATCACTTA  | AAAGGTGTTA  | TAAGTAAAT   | ATCGGTAATA  | AAAGGTGGCC  | CAAAGTGAAA  | TTTACTCTTT  |
| 4271 | TCTACTATTA  | TTAAAATTGA  | GGATGTTTTG  | TGGTACTTT   | GATAACGTCA  | TTTTGTATGA  | ATTGGTTTT   |
| 4341 | AAGTTTATTTC | GGGATTGTTGA | AATGGCATATC | TGTATTGAG   | TGGTTTTTA   | AGTTCGTTGC  | TTTGTAAAT   |
| 4411 | ACAGAGGGAT  | TGTATAAGA   | AATATCTTA   | AAAACCCAT   | ATGCTTAATT  | GACATAATT   | TTGAGAAAAA  |
|      |             |             |             |             | ECORI       | ~~~~~       |             |
| 4481 | TATATATTCA  | GGCGAATTCC  | ACAATGAACA  | ATAATAAGAT  | TTAAATAGCT  | TGCCCGCGT   | GCAGCGATGG  |
| 4551 | GTATTTTTC   | TAGTAAATA   | AAAGATAAAC  | TTAGACTCAA  | AACATTACA   | AAAACAACCC  | CTAAAGTCCCT |
| 4621 | AAAGCCAAA   | GTGCTATGCA  | CGATCCATAG  | CAAGGCCAGC  | CCAACCCAAAC | CCACCCCAAC  | CCACCCAGT   |
| 4691 | GCAGCCAACT  | GGCAAATAGT  | CTCACCCTCC  | GGCACATATCA | CCGTGAGTTG  | TCCGCACAC   | CGCACGCTCTC |
| 4761 | GCAGCCAAA   | AAAAAAAG    | AAAGAAAAAA  | AGAAAAGA    | AAACAGGCAG  | GTGGGTCCGG  | GTCGTGGGGG  |
| 4831 | CCGGAAAAGC  | GAAGGAGGATC | GCGAGCAGCG  | ACGAGGCCCG  | GCCCTCCCTC  | CGCTTCCAAA  | AAAACGCC    |

FIG.\_33D

|       |            |             |            |            |            |            |            |
|-------|------------|-------------|------------|------------|------------|------------|------------|
| 44901 | CCATGCCAC  | TATATACATA  | CCCCCCCTC  | TCCTCCCATC | CCCCCAACCC | TACCAACC   | ACCACCA    |
| 44971 | CCTCCTCCC  | CCTCGTGCC   | GGACGACGAG | CTCCCTCCCC | CTCCCCCTCC | GCGGCCGCG  | GTAACCACCC |
| 5041  | CGCCCTCTC  | CTCTTCTTT   | CTCCGTTTT  | TTTTTCGTCT | CGGTCTCGAT | CTTGGCCCTT | GGTAGTTGG  |
| 5111  | GTGGCGAGA  | GCGGCTTCGT  | CGCCCAGATC | GGTGGGGGG  | AGGGGGGGGA | TCTCGGGGT  | GGCGTCCTCG |
|       |            |             |            |            |            | BglII      |            |
|       |            |             |            |            |            |            | ~~~~~      |
|       |            |             |            |            |            | BamHI      | ~~~~~      |
| 5181  | GGGGTGAGTC | GGCCCCGGATC | CTCGGGGGGA | ATGGGGCTCT | CCGATGTA   | TCTTCCTTCT | TTCCTCTTTT |
| 5251  | TGTGGTAGAA | TTTGAATCCC  | TCAGCATTGT | TCATCGGTAG | TTTTTCTTT  | CATGATTGT  | GACAAATGCA |
| 5321  | GCCTCGTGGC | GAGCTTTTT   | GTAGC      |            |            |            |            |

## FIG.\_33E



**FIG.\_34A**

NcoI

HindIII

1 AAGCTTACCA TGGCCCACGC CCGGTCTTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCGTGCACG

Not I

• A S S S F A D S N P I R P V T D R A A S T .  
71 TCGCTCCTC CTCCTCCTC GCCGACTCCA ACCCGATCCG GCCGTCACC GACCGCGGG CGCCCTCCAC  
• Q G I S E D L Y S R L V E M A T I S Q A A Y A  
141 GCAGGGCATC TCCGAAGAACCTACAGGCC TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCCTACGCC

SalI

• D L C N I P S T I I K G E K I Y N S Q T D I N G  
211 GACCTGTGCA ACATTCCGTC GACTATTAC AAGGGAGAGA AAATTACAA TTCTCAAAT GACATTAACG

BamHI

• W I L R D D S S K E I I T V F R G T G S D T N .  
281 GATGGATCCT CGCGGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCGCT GGCACCTGGTA GTGATACGAA  
• L Q L D T N Y T L T P F D T L P Q C N G C E V  
351 TCTACACTC GATACTAATC ACACCCCTCAC GCCTTTCGAC ACCCTTACAC AATGCAACGG TTGTGAAGTTA  
H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S  
421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTGGAGTC GCTTGTCAA CAGCAGGTTA  
• Q Y P D Y A L T V T G H X L G A S L A A L T A .  
491 GCCAGTATCC GGACTACCGG CTGACCGTGA CGGGCCACKC CCTCGGGCC TCCCTGGGG CACTCACTGC  
• A Q L S A T Y D N I R L Y T F G E P R S G N Q  
561 CGCCCAAGCTG TCTGGACAT AGGACAACAT CGGCTGTAC ACCTTCGGCG AACCCGGCAG CGGCAATCAG

XbaI

• A F A S Y M N D A F Q A S S P D T T Q Y F R V T  
631 GCCTTCGGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATA GACGCAGTAT TTCCGGTCA

NcoI

• H A N D G I P N L P P V E Q G Y A H G G V E Y .

FIG.-34B

|     |                                                                               |
|-----|-------------------------------------------------------------------------------|
| 701 | CTCATGCCAA CGACGGCATC CCAAACCTGC CCCGGTGGAA GCAGGGTAC GCCCATGGCG GTGTTAGAGTA  |
| 702 | • W S V D P Y S A Q N T F V C T G D E V Q C C E                               |
| 771 | CTGGAGCGTT GATCCTTACA GGGCCCAGAA CACATTGTC TGCACTGGGG ATGAAAGTGGCA GTGCTGTGAG |
|     | SphI                                                                          |

841 GCCCAAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTGTTGGAT GACGAGGGCC GCATGGACCT

Moti

|                                                                                                                                           |                                                                                                               |                                                                                                      |
|-------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|
| . P V A A E P L K D E L *                                                                                                                 | . P V A A E P L K D E L *                                                                                     | . P V A A E P L K D E L *                                                                            |
| 911 GGC CGGT CGC GGG CGGG A CCA CTG AAG G ATG AGCT GT A AGA AGC AGA TCG TTCA AAC ATT GGC AAT                                              | 981 AAAG TTCT T AAG ATT GAA T GCA T GAT GCG AT GCT GCG G GTCT TGCG C GAT TAT CCA TA ATT TCT GT TGA ATT TAC GT | 1051 TAGG CAT GT A ATA ATT A A C A T G C A T G C T T A T G G G T T T T T A T G A T T A G G T C C C G |
| 1121 CAA TT A C A T T T A A T A C G C G A T A G A A A A C C G C G C A A A C T A G G A T A A A T T A G G A T A A A T T A T C G G G C G G G |                                                                                                               |                                                                                                      |

Hindustan

અનુભાવ

|       |             |             |             |            |            |             |             |
|-------|-------------|-------------|-------------|------------|------------|-------------|-------------|
| 11191 | TGTTCATCTAT | GTTACTAGAT  | CGATAAGCTT  | CTAGAGCCC  | CGCTGGAGCT | CCAATTGCC   | CTATAGTGAG  |
| 11261 | TGGTATTACG  | CGGGCTCACT  | GGCGGTGTT   | TTACAACGTC | GTGACTGGGA | AAACCCCTGGC | GTTACCCAAAC |
| 11331 | TTAATCGCCT  | TGCAGGCACAT | CCCCCTTTCG  | CCAGCTGGCG | TAATAGCGAA | GAGGCCCGCA  | CCGATCGCCC  |
| 11401 | TTCCCAACAG  | TTGCCAGGCC  | TGAATGGCGA  | ATGGGACGGG | CCCTGTAGCG | GGCATTAAAG  | CGGGGGGGT   |
| 11471 | GTGGGGTTA   | CGCGCAGCGT  | GACCGCTACA  | CTTGGCAGCG | CCCTAGCGCC | CGCTCCCTTC  | GCTTTCTTCC  |
| 11541 | CTTCCTTCTC  | CGCCACGTTC  | GGGGCTTTC   | CCCAGTCAGC | TCTAAATCGG | GGGCTCCCT   | TAGGGTCCG   |
| 11611 | ATTAGTGCT   | TTACGGCAC   | TCGACCCCCA  | AAAACCTGAT | TAGGGTGTAG | GTTCACGTAG  | TGGGCCATCG  |
| 11681 | CCCTGATAAG  | CGGTTTTCG   | CCCTTGACG   | TTGGAGTCCA | CGTTCTTAA  | TAGTGGACTC  | TTGTTCCAAA  |
| 11751 | CTGGAAACAC  | ACTCAACCT   | ATCTGGTCT   | ATTCTTTGA  | TTATAAGGG  | ATTGGCCGA   | TTTGGCTA    |
| 11821 | TTGGTTAAA   | AATGAGCTGA  | TTAACAAA    | ATTAACGGC  | AAATTAAACA | GCTTACAATT  |             |
| 11891 | TAGGTGGCAC  | TTTTCGGGG   | AATGTGGCG   | GAACCCCTAT | TGTTTATTT  | TTCTAAATAC  | ATTCAAATAT  |
| 11961 | GTATCCGGCTC | ATGAGACAAT  | AACCCGTATA  | AATGCTCAA  | TAATATGAA  | AAAGGAAGAG  | TATGAGTATT  |
| 2031  | CAACATTTC   | GTGTGGCCCT  | TATTCCTTT   | TTTGGGCCAT | TTGCTTGGCT | TGTTTTGGCT  | CACCCAGAAA  |
| 2101  | CGCTGGTGA   | AGTAAAGAT   | GCTGAAGATC  | AGTTGGTGC  | ACGAGTGGGT | TACATCGAAC  | TGGATCTCAA  |
| 21171 | CAGGGTAAAG  | ATCCCTGAGA  | GTTCGCCCC   | CGAAGAACGT | TTTCCAATGA | TGAGCAGTT   | TAAGTTCTG   |
| 22241 | CTATGTTGGC  | CGGTATTATC  | CCGTATTGAC  | GGCGGGCAAG | AGCAACTCGG | TGCCCCGCTA  | CACTATTCTC  |
| 23111 | AGAATGACTT  | GGTGTAGTAC  | TCACCAAGTCA | CAGAAAGCA  | TCTTACGGAT | GGCATGACAG  | TAAGGAATT   |
| 23811 | ATGCAGTGGCT | GGCATTAACCA | TGAGTGTATAA | CAGAAAGCA  | AACTTACTTC | TGACAAACGAT | CGGAGGACCG  |

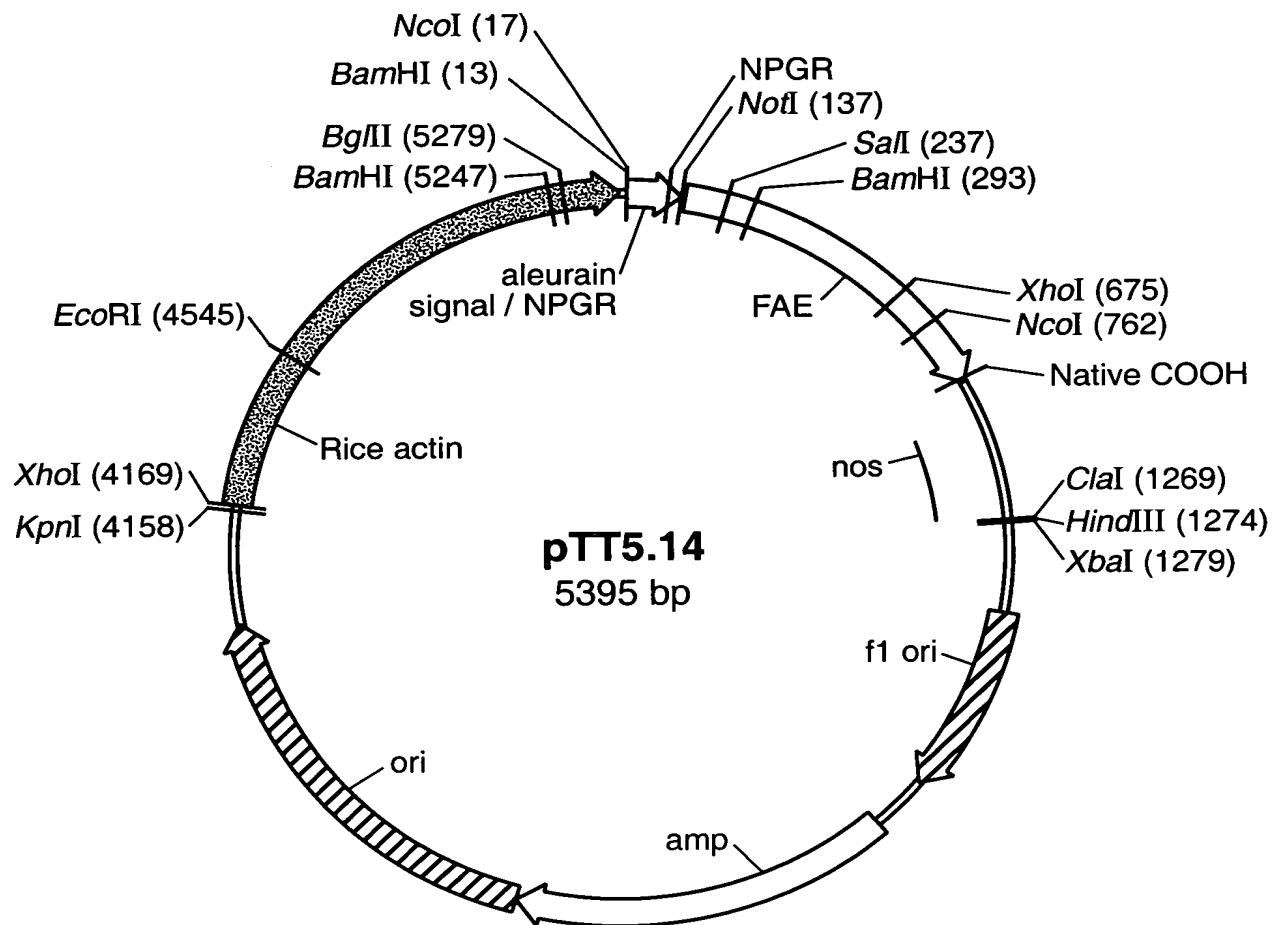
FIG. 34C

|      |             |             |             |             |             |             |             |
|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 2451 | AAGGAGCTAA  | CCGCTTTTT   | GCACAACATG  | GGGGATCATG  | TAACTCGCCT  | TGATCGTTGG  | GAACCGGGAGC |
| 2521 | TGAATGAGC   | CATAACAAAC  | GACGAGCGTG  | ACACCACTG   | GCCTGTAGCA  | ATGGCAACAA  | CGTTGGCAA   |
| 2591 | ACTATTAAC   | GGCGAACTAC  | TTACTCTAGC  | TTCCCAGCAA  | CAATTAATAG  | ACTGGATGGA  | GGGGATAAA   |
| 2661 | GTTCGAGGAC  | CACTTCTGG   | CTCAGCCCTT  | CCGGCTGGCT  | GGTTTATTGC  | TGATAAATCT  | GGAGCCGGTG  |
| 2731 | AGCGTGGTC   | TCGGCGGTATC | ATTGCAGCAC  | TGGGCCAGA   | TGGTAAGCCC  | TCCCCTATCG  | TAGTTATCTA  |
| 2801 | CACGACGGGG  | AGTCAGGCAA  | CTATGGATGA  | ACGAAATAGA  | CAGATCGCTG  | AGATAAGGTGC | CTCACTGATT  |
| 2871 | AAGCATGGT   | AACTGTCAGA  | CCAGTGTAC   | TCATATATAC  | TTAGATTGA   | TTAAAAACTT  | CATTTTTAAAT |
| 2941 | TTAAAGGAT   | CTAGGTGAAG  | ATCCTTTTG   | ATAATCTCAT  | GACCAAAATC  | CCTTAACGTG  | AGTTTTCGTT  |
| 3011 | CCACTGAGCG  | TCAGACCCCG  | TAGAAAAGAT  | CAAAGGATCT  | TCTTGAGATC  | CTTTTTTCT   | GGCGGTAATC  |
| 3081 | TGCTGCTTGC  | AAACAAAAAA  | ACCAACGGTA  | CCAGGGTGG   | TTTGTGTTGCC | GGATCAAGAG  | CTACCAACTC  |
| 3151 | TTTTTCGAA   | GGTAACTGGC  | TTAGCAGAG   | CGCAGATAAC  | AAATAACTGTG | CTTCTAGTGT  | AGCCGTAGTT  |
| 3221 | AGGCCACCAC  | TTCAGAACT   | CTGTAGCAC   | GCCTACATAC  | CTCGCTCTGC  | TAATCCTGTT  | ACCAAGTGGCT |
| 3291 | GCTGCCAGTG  | GGGATAAGTC  | GTGTCTTAC   | GGGTGGACT   | CAAGACGATA  | GTACCGGGAT  | AAGGGCGAGC  |
| 3361 | GGTGGGGCTG  | AACGGGGGT   | TGTCACAC    | AGCCCAGCTT  | GGAGCGAACG  | ACCTACACCG  | AACTGAGATA  |
| 3431 | CCTACAGCGT  | GAGCTATGAG  | AAAGGCCAC   | GCTTCCGAA   | GGAGAAAGG   | CGGACAGGTA  | TCCGGTAAGC  |
| 3501 | GGCAGGGTCG  | GAACAGGAGA  | GCGCACGAGG  | GAGCTTCCAG  | GGGAAACGC   | CTGGTATCTT  | TATAGTCCGT  |
| 3571 | TGGGTTTCG   | CCACCTCTGA  | TTGAGGTC    | GATTTTGTG   | ATGCTCGTCA  | GGGGGGGGA   | GCCTATGGAA  |
| 3641 | AAAGCCAGC   | AACGGGGCCT  | TTTACGGT    | CCTGGCCTT   | TGCTGGCCTT  | TTGCTCACAT  | GTCTTCTCCT  |
| 3711 | GGGTATCCC   | CTGATTCTGT  | GGATAACCGT  | ATTACCGCT   | TTGAGTGGAC  | TGATAACGGCT | CGCCGGAGCC  |
| 3781 | GAACGACCGA  | GCGCAGCGAG  | TCAGTGAGG   | AGGAAGGGGA  | AGAGGGCCCA  | ATACGAAAC   | CGCCCTCTCCC |
| 3851 | CGCGGTTGG   | CCGATTCA    | ATAGGACGTG  | GCACGACAGG  | TTTCCCGACT  | GGAAAGGGG   | CAGTGAGGCC  |
| 3921 | AACGCAATT   | ATGTGAGTTA  | GCTCACTCAT  | AGGCACCCC   | AGGCTTACA   | CTTTATGCTT  | CGGGCTCGTA  |
| 3991 | TGTTGTGG    | AATTGTGAGC  | GGATAACAAAT | TTCACACAGG  | AAACAGCTAT  | GACCATGATT  | ACGCCAAGGG  |
|      |             |             |             |             | XbaI        | KpnI        |             |
| 4061 | CGCAATTAAAC | CCTCACTAAA  | GGGAACAAAAA | GCTGGGTACC  | GGGGCCCCCC  | TCGAGGTCT   | TCATATGCTT  |
| 4131 | GAGAAGGAG   | TGGGATAGT   | CCAAAATAAA  | ACAAGGTA    | GATTACCTGG  | TCAAAAGTGA  | AAACATCAGT  |
| 4201 | TAAAAGTGG   | TATAAGTAA   | ATATCGGTA   | TAAAAGTGG   | CCAAAGTGA   | AAATTACTCT  | TTTCTACTAT  |
| 4271 | TATAAAATT   | GAGGTGTTACT | TGTGGTACT   | TTGATACGTC  | ATTTTGTAT   | GAATTGGTTT  | TTAAGTTTAT  |
| 4341 | TCGGGATTTC  | GAAATGCCATA | TCTGTATTG   | AGTCGGTTT   | TAAGTTGTT   | GCTTTTGAA   | ATACAGAGGG  |
| 4411 | ATTGTATAA   | GAATAATCTT  | AAAAAACCC   | ATATGCTAA   | TTGACATAAT  | TTTTGAGAAA  | AATATATAAT  |
|      |             |             |             |             | ECORI       |             |             |
| 4481 | CAGGGAAATT  | CCACAATGAA  | CAATAATAAG  | ATTAATAATAG | CTTGCCCCCG  | TTGCAGCGAT  | GGGTATTTT   |
| 4551 | TCTAGTAAA   | AAAAAGATAAA | ACTTAGACTC  | AAAACATTA   | CAAAACAAAC  | CCCTAAAGTC  | CTAAAGCCCA  |

FIG.-34D

4621 AAGTGCTATG CACGGATCCAT AGCAAGCCCCA GCCCCAACCCA ACCCAACCCCA GTGCAGCCAA  
4691 CTGGCCAATA GTCTCCACCC CCGGGCACTAT CACCGTGAGT TGTCGGCACC ACCGGCACGTC TCGCAGCCAA  
4761 AAAAGAAAAA AGAAAGAAAAA AAAAGAAAAA GAAAACAGC AGGTGGGTCC GGGTCGTGGG GGCGGGAAA  
4831 GCGAGGGAGGA TCGGGAGCGAG CGACGGAGGCC CGGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC  
4901 ACTATATACA TACCCCCCCC TCTTCTCTCCA TCCCCCAAC CCTAACCAAC CCACCAAC CACCTCCTCC  
4971 CCCCTCGCTG CCGGACGAGC AGCTCCCTCCC CCCTCCCCCT CGGCTGGCC CGGTAAACAC CCCGGCCCTC  
5041 TCCTCTTTCT TTCTCCGTTT TTTTTCTCGT ATCTTGGCC ATCTTGGCC TTTGGTAGTT GGGTGGGGA  
5111 GAGGGGCTTC GTCGCCCCAGA TCGGTGGGG GGAGGGGGG GATCTCGGGG CTGGCGTTC CGGGCGTCA  
BamHI ~~~~~  
5181 TCGGCCCGGA TCCTCGGGG GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTTGTGGTAG  
5251 AATTGAAATC CCTCAGGATT GTTCATCGGT AGTTTTCTT TTCAATGTT GTGACAAATG CAGGCTCGTG  
5321 CGGAGCTTTT TTGTAGC

FIG.\_34E



**FIG.\_35A**

~~~~~ NcoI ~~~~~  
~~~~~ BamHI ~~~~~

1 CCTGACGCCG AGGATCCATG GCCCACGCC GCGTCCCTCCT CCTGGGGCTC GCCGGTGG CCACGGCCGC  
Not I

71 CGTCGCCGTC GCCTCCTCT CCTCCTCGC CGACTCCAAC CGGGCCGGC CGGTACCGA CGCGGGCC

~~~~~ A S T Q G I S E D L Y S R L V E M A T I S Q A A  
141 GCCTCCACGC AGGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAT GGCCACTATC TCCAAAGCTGA
Sali

~~~~~ Y A D L C N I P S T I I K G E K I Y N S Q T D  
211 CCTACGCCGA CCTGTGCAAC ATTCCGTGCA CTATTATCAA GGGAGAGAAA ATTACAATT CTCAAACTGA  
BamHI

~~~~~ I N G W I L R D D S S K E I I T V F R G T G S  
281 CATTAAAGGA TGGATCCCTCC GCGACGGACAG CAGGAAAGAA ATAATTCACCG TCTTCGGTGG CACTGGTAGT
D T N L Q L D T N Y T L T P F D T L P Q C N G C

351 GATACTGAAATC TACAACCTCGA TACTAACCTAC ACCCCTCACCGC CTTTCGACAC CCTTACACAA TGCAACGGTT
BamHI

~~~~~ E V H G G Y Y I G W V S V Q D Q V E S L V K Q  
421 GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA GTCGAGTCGC TTGTCAAACA  
Q V S Q Y P D Y A L T V T G H X L G A S L A A

491 GCAGGGTTAGC CAGTATCCGG ACTACGGGT GACCGTGACC GGCCACKCCC TCGGGCGCTC CCTGGGGCA  
L T A A Q L S A T Y D N I R L Y T F G E P R S G

561 CTCACTGCCG CCCAGCTGTC TGCGACATAC GACAACATCC GCCTGTACAC CTTCGGGAA CGCGCAGCG  
XbaI

~~~~~ N Q A F A S Y M N D A F Q A S S P D T T Q Y F  
631 GCAATCAGGGC CTTCGCGTCG TACATGAACG ATGCCTCCA AGCCTCGAGC CCAGATAACGA CGCAGTATT
NcoI

~~~~~ R V T H A N D G I P N L P P V E Q G Y A H G G

FIG.-35B

|      |                                                                                       |
|------|---------------------------------------------------------------------------------------|
| 701  | CCGGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCC CCGGTGGAC AGGGTACGC CCATGGGGT              |
|      | V E Y W S V D P Y S A Q N T F V C T G D E V Q C                                       |
| 771  | GTAGAGTACT GGAGC GTTGA TCCTTACAGC GCCCAGAAC A CATTGCTG CACTGGGAT GAAGTGCAGT           |
|      | · C E A Q G G Q G V N N A H T T Y F G M T S G A ·                                     |
| 841  | GCTGTGAGGC CCAGGGCGGA CAGGGCGGA ATAATGCCA CACGACTAT TTGGGTATGA CGAGCGGAGC             |
|      | · C T W *                                                                             |
| 911  | CTGTACATGG TGATCAGTCA TTTTCAGCCTC CCCGAGTGTAA CCAGGAAAGA TGGATGTCTCT GGAGAGGGGG       |
| 981  | CCGGTAAACC ACTGAAGGAT GAGCTGTAAA GAAGCAGATC GTTCAAACAT TTGGCAATAA AGTTTCTTAA          |
| 1051 | GATTGAATCC TGTGCGGGT CTTGCGATGA TTATCATATA ATTCTGTG AATTACGTTA AGCATGTTA              |
| 1121 | AATTAAACATG TAATGCCATGA CGTTATTAT GAGATGGGTT TTATGATTAA GAGTCCCGCA ATTACATT           |
| 1191 | TAATACGCGA TAGAAACCAA AATATAGCGC GCAAACCTAGG ATAAATTATC GCGGCCGGTG TCATCTATGT HindIII |

|      |             |             |             |             |             |             |             |
|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1261 | TACTAGATCG  | ATAAGCTTCT  | AGAGGGCCG   | GTTGGAGCTCC | ATTGCCCT    | ATAGTGAGTC  | GTATAGCGC   |
| 1331 | CGCTCACTGG  | CCGTCGTTT   | ACAACGTCGT  | GAATGGAAA   | ACCCCTGGCGT | TACCCAACCT  | AATCGCCTTG  |
| 1401 | CAGCACATCC  | CCC'TTTGCCC | AGCTGGCGTA  | ATAGGAAAGA  | GCCCCGGACC  | GATGCCCTT   | CCCAACAGTT  |
| 1471 | GCGCAGCCCTG | AATGGCGAAT  | GGGACGGGCC  | CTGTAGGGGC  | GCATTAAGGC  | CGGGGGGTGT  | GGTGGTTACG  |
| 1541 | CGCAGCGTGA  | CCGCTTACACT | TGCCAGGGCC  | CTAGGGCCCG  | CTCCCTTTCGC | TCTTCTCCCT  | TCCTTCTCG   |
| 1611 | CCACGTTGCC  | CGGCTTCCC   | CGTCAAGCTC  | TAATCGGGG   | GCTCCCTTAA  | GGGTTCCGAT  | TTAGTGCTTT  |
| 1681 | ACGGCACCTC  | GACCCCCAAA  | AACTTGATTA  | GGGTGATGGT  | TCACGTTAGTG | GGCCTATCGGC | CTGATAGACG  |
| 1751 | GTTTTTCGCC  | CTTGACGTT   | GGAGTCCACG  | TTCTTTAATA  | GTGGACTCTT  | GTTCACAAACT | GGAACAAACAC |
| 1821 | TCAAACCTAT  | CTCGGTCTAT  | TCTTTTGATT  | TATAAGGGAT  | TTGGCCGATT  | TGCGCTTAT   | GGTTAAAAAA  |
| 1891 | TGAGCTGATT  | TAACAAAAAT  | TAAACGGAA   | TTTTAACAAA  | ATATTAACGC  | TTACAATTTA  | GGTGGCACTT  |
| 1961 | TTGGGGAAA   | TGTGGGGGA   | ACCCCTATT   | GTTTATTTTT  | CTAAATACAT  | TCAAATATGT  | ATCCGCTCAT  |
| 2031 | GAGACAATAA  | CCCTGATAAA  | TGCTTCAATA  | ATATTGAAAAA | AGGAAGAGTA  | TGAGTATTCA  | ACATTTCGGT  |
| 2101 | GTGCCCTTA   | TTCCCTTTT   | TGGGGCATTT  | TGGCTTCCCTG | T'TTTGCTCA  | CCCAGAAAAG  | CTGGTGAAAG  |
| 2171 | TAAGAGATGC  | TGAAGATCAG  | TTGGGTGCAC  | GAGTGGGTTA  | CATCGAACTG  | GCGGTAAGAT  |             |
| 2241 | CCTTGAGAGT  | TTTGCCCCCG  | AAGAACGTT   | TCCAATGATG  | AGCACTTTA   | AAGTTCTGCT  | ATGGGGCGCG  |
| 2311 | GTATTATCCC  | GTATTGACGC  | CGGGCAAGAG  | CAACTCGGTC  | GCCGGCATACA | CTATTCTCAG  | AATGACTTGG  |
| 2381 | TGAGTACTC   | ACCAAGTCACA | GAAAGGCATC  | TTACGGATGG  | CATGACAGTA  | AGAGAAATT   | GCAGTGCTGC  |
| 2451 | CATAACCATG  | AGTGATAAAC  | CTGGGGCAA   | CTTACTTCTG  | ACAACGATCG  | GAGGACGAA   | GGAGCTAAC   |
| 2521 | GCTTTTTGTC  | ACAAACATGGG | GGATCATGTA  | ACTCGGCCTTG | ATCGTTGGGA  | ACCGGAGCTG  | AATGAAGGCCA |
| 2591 | TACCAAAACGA | CGAGCGTGAC  | ACCAAGATGC  | CTGTAGCAAT  | GGCAACACAG  | TTGCGCAAAC  | TATTAACCTGG |
| 2661 | CGAAACTCTT  | ACTCTAGCTT  | CCCCGGCAACA | ATTAATAGAC  | TEGGATGGAGG | CGGGATAAAGT | TGCGGAGACCA |

FIG. 35C

|       |             |             |            |             |             |              |             |
|-------|-------------|-------------|------------|-------------|-------------|--------------|-------------|
| 27731 | CTCTCGGCT   | CGGCCCTCC   | GGCTGGCTGG | TTTATTCTG   | ATAAATCTG   | GGCCGGTGTGAG | CGTCGGGTCTC |
| 27801 | GCGGTATCAT  | TGCAAGCACTG | GGGCCAGATG | GTAAGCCCTC  | CCGTATCGTA  | GTTATCTACA   | CGACGGGGAG  |
| 27871 | TCAGGCAACT  | ATGGATGAAC  | GAAATAGACA | GATCGCTGAG  | ATAGGTGCCT  | CACTGATTAA   | GCATGGTGTAA |
| 27941 | CTGTCAGACC  | AAGTTTACTC  | ATATATACTT | TAGATTGATT  | TAAAACCTCA  | TTTTTAATT    | AAAGGGATCT  |
| 30011 | AGGTGAAGAT  | CCTTTTGTAT  | AATCTCATGA | CCAAAATCCC  | TTAACGTGAG  | TTTCGTTTC    | ACTGAGGGTC  |
| 30081 | AGACCCCGTA  | GAAAAGATCA  | AAGGATCTTC | TTGAGATCCT  | TTTTTTCTGC  | GCGTAATCTG   | CTGCTTGGAA  |
| 31511 | ACAAAAAAAC  | CACCGCTAAC  | AGGGGTGTTT | TGTTTGCCTG  | ATCAAGAGCT  | ACCAACTCTT   | TTTCCGAAGG  |
| 31521 | TAACTGGCTT  | CAGGAGGCG   | CAGATACCAA | ATACTGTCCT  | TCTAGTGTAG  | CCGTAAGTTAG  | GCCACCAACTT |
| 32911 | CAAGAACTCT  | GTAGCACCGC  | CTACATACCT | CGCTCTGCTA  | ATCCTGTAC   | CAGTGGCTGC   | TGCCIACTGGC |
| 33361 | GATAAGTCTG  | GTCTTACCGG  | GTGGAECTCA | AGACGATAGT  | TACCGGATAA  | GGCCGAGCGG   | TGGGGCTGAA  |
| 33431 | CGGGGGTTC   | GTGCAACACAG | CCCAGCTGG  | CTACACGGAA  | CTGAGATAACC | TACAGCGTGA   |             |
| 33501 | GCTATGAGAA  | AGGCCACAGC  | TTCGGAAAGG | GAGAAAGGGC  | GACAGGTATC  | CGGGTCCGG    |             |
| 33571 | ACAGGAGAC   | GCACGAGGA   | GCTTCAGGG  | GGAAACCCCT  | GGTATCTTA   | TAGTCCTGTC   | GGGTTTCGCC  |
| 33641 | ACCTCTGACT  | TGAGCGTCA   | TTTTTGTGAT | GCTCGTCA    | GGGGGGAGG   | CTATGGAAAA   | ACGCCAGCAA  |
| 33711 | CGGGGCCTT   | TTACGGTTCC  | TGGCCTTTG  | CTGGCCTTT   | GCTCACATGT  | TCTTTCCTGC   | GTТАТССССТ  |
| 33781 | GATTCTGTGG  | ATAACCGTAT  | TACCGCCTT  | GAGTGAGCTG  | ATACCGTCTG  | CCGAGGCCGA   | ACGACCGAGC  |
| 33851 | GCAGCGAGTC  | AGTGAGCGAG  | GAAGGGGAAG | AGGGGCCAAT  | ACGCAAACCG  | CCTCTCCCCG   | CGGGTTGGCC  |
| 33921 | GATTCAATTAA | TGCAAGTGGC  | ACGACAGGT  | TCCCCGACTGG | AAAGGGGGCA  | GTGAGCGCAA   | CGCAATTAAAT |
| 33991 | GTGAGTTAGC  | TCACTCAATT  | GGCACCCAG  | GCTTTACACT  | TTATGCTTCC  | GGCTCGTATG   | TGTGTGGAA   |
| 40611 | TTGTGAGCGG  | ATAACCAATT  | CACACAGAA  | ACAGCTATGA  | CCATGATTAC  | GCCAAAGGGCG  | CAATTAAACCC |

|       | KpnI       | XbaI        |             |
|-------|------------|-------------|-------------|
| 44131 | TCACAAAGG  | GAACAAAAGC  | TGGGTACCGG  |
| 44201 | GGGATAGTC  | AAAATAAAC   | AAAGGTAAGA  |
| 44271 | TAAGTAAAT  | ATCGGTAAATA | AAAGGTGGCC  |
| 44341 | GGATGTTTG  | TCGGTACTTT  | GATACTCTAT  |
| 44411 | AATGCATATC | TGTATTTGAG  | TCGGTTTTTA  |
| 44481 | AATATCTTA  | AAAAACCCAT  | ATGCTAATT   |
| 44551 | ACCATGACA  | ATATAAGAT   | TAAATAGCT   |
| 44621 | AAAGATAAAC | TTAGACTCAA  | AAACATTACA  |
| 44691 | CGATCCATAG | CAAGCCCAGC  | CCAACCCAAC  |
| 44761 | CTCCACCCCC | GGCACTATCA  | CCGTGAAGTTG |
| 44831 | AAAGAAAAAA | AAGIAAAGA   | AAACAGCAG   |

FIG. 35D

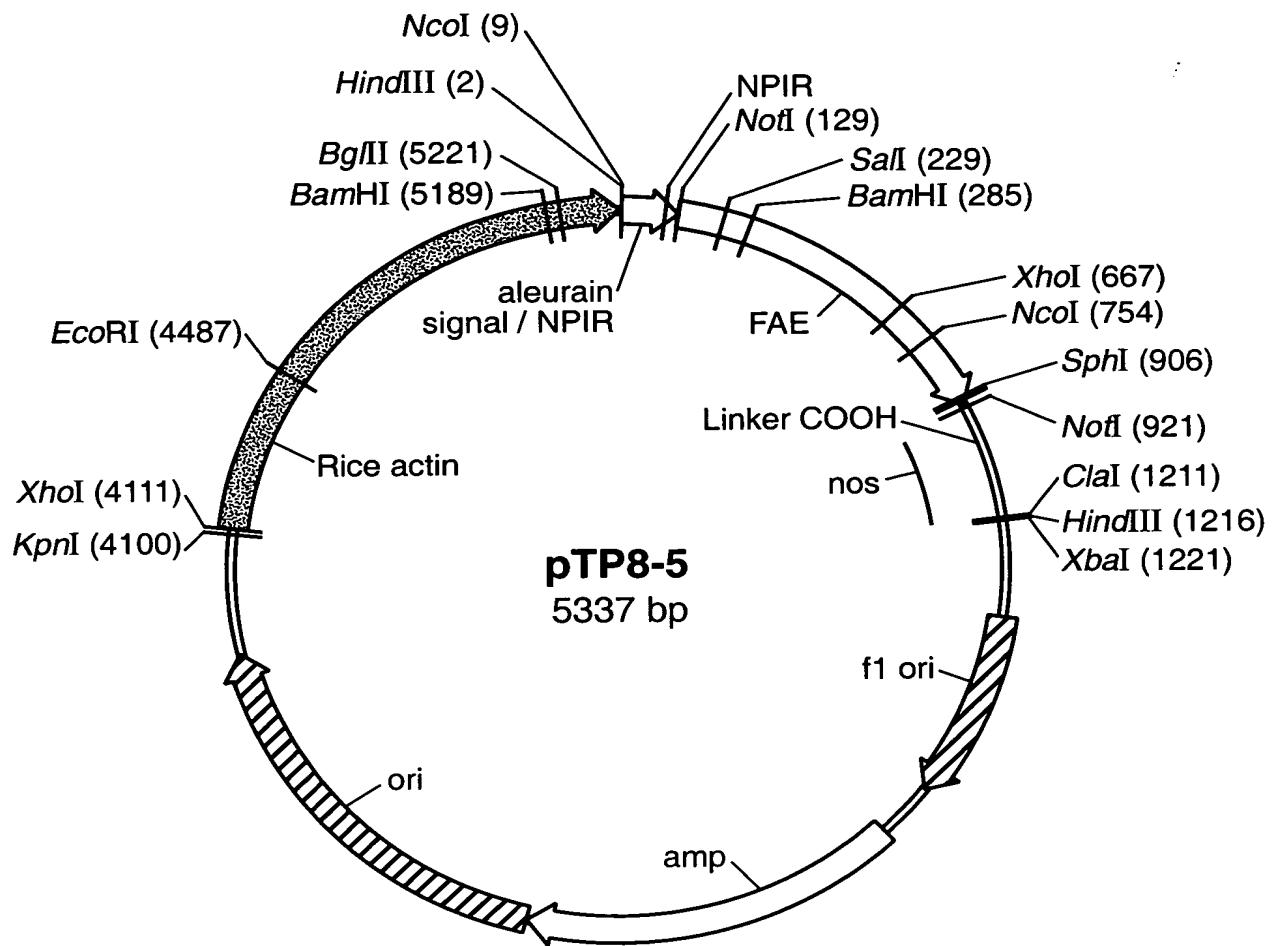
4901 GCGAGGAGCG ACAGGGCCCG GCCCTCCCTC CGCTTCCAA GAAACGCCCG CCATGCCAC TATATACATA  
4971 CCCCCCCCTC TCCTCCCATC CCCCAACCC TACCACCCACC ACCACCAACCA CCTCCCTCCCC CCTCGCTGCC  
5041 GGACGACGAG CTCCCTCCCG CTCCCTCCCG GCGGCCGCCG GTAACCACCC CGCCCCCTCTC CTCCTTCTT  
5111 CTCCGTTTTT TTTCGTTCT CGGTCTCGAT CTTCGGCCTT GGTAGTTGG GTGGCGAGA GGGGCTTCGTT  
BamHI

5181 CGCCCAAGATC GGTGCGCGGG AGGGGGGGGA TCTCGCGGCT GGCGTCTCCG GGGGTGAGTC GGCCCCGGATC  
BamHI

~ ~~~~~

5251 CTCGCGGGGA ATGGGGCTCT CGGATGTAGA TCTTCTCTCTCT TTCTTCTCTTT TGTGGTAGAA TTTGAATCCC  
5321 TCAGGCATTGT TCATCGGTAG TTTTCTTTT CATGATTGT GACAAATGCA GCCTCGTGC GAGGCTTTTT  
5391 GTAGC

*FIG.\_35E*



**FIG.\_36A**

NcoI

HindIII

~~~~~  
M A H A R V L L A L A V L A T A A V A V
1 AAGCTTACCA TGGCCACGC CCGGTCTC CTCCTGGCGC TCGCCGTGCT GCCCACGGCC GCCGTGCCG
NotI

~~~~~  
A S S F A D S N P I R P V T D R A A S T  
71 TCGCCTCCTC CTCCCTCCTC GCCGACTCCA ACCCGATCCG GCCGTCTCACC GACCGCGGG CGGCCTCCAC  
Q G I S E D L Y S R L V E M A T I S Q A A Y A  
141 GCAGGGCATC TCCGAAGACC TCTACAGCCC TTAGTCGAA ATGGCCACTA TCTCCAAAGC TGCCTACGCGC

SalI

~~~~~  
D L C N I P S T I I K G E K I Y N S Q T D I N G
211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTACAA TTCTCAAACT GACATTAACG
BamHI

BamHI

~~~~~  
W I L R D D S S K E I I T V F R G T G S D T N  
281 GATGGATCCT CGCGACGAC AGCAGCAAG AAATAATCAC CGTCTTCCGT GGCACGTGTA GTGATACGAA  
L Q L D T N Y T L T P F D T L P Q C N G C E V  
351 TCTACAACTC GATACTAAT ACACCCCTCAC GCCTTTCGAC ACCCTTACAC AATGCAACGG TTGTGAAGTA  
H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S  
421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA  
Q Y P D Y A L T V T G H X L G A S L A A L T A  
491 GCCAGTATCC GGACTACGGG CTGACCGTGA CGGGCCACKC CCTCGGGCC TCCCTGGGG CACTCACTGG  
A Q L S A T Y D N I R L Y T F G E P R S G N Q  
561 CGCCCAAGCTG TCTGGACAT ACCACAACAT CGGCCGTGAC ACCTTTCGGGG AACCGGCCAG CGGCAATCAG  
XbaI

~~~~~  
A F A S Y M N D A F Q A S S P D T T Q Y F R V T
631 GCCTTCGGGT CGTACATGAA CGATGCCTTC CAAGCCCTCGA GCCCAGATAAC GACGGCAGTAT TTCCGGGTCA
NcoI

~~~~~  
H A N D G I P N L P P V E Q G Y A H G G V E Y .

FIG.\_36B

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCGGGTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
• W S V D P Y S A Q N T F V C T G D E V Q C C E  
771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTTGTC TGCACGGGGG ATGAAAGTGC GTGCTGTGAG  
SphI

841 A Q G G Q G V N N A H T T Y F G M T S G A C T W  
GCCCAAGGGGG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGAT GACGAGGGGC GCATGGCACCT

Not I

• P V A A A \*  
911 GGCGGGTGC GGGCGGTAA CCACTGAAGG ATGAGCTGTA AAGAACAGA TCGTTCAAAAC ATTGGCAAT  
981 AAAGTTCTT AAGATTGAAT CCTGTTGCCG GTCTTGGAT GATTATCATA TAATTCTCTGT TGAATTACGT  
1051 TAAGCATGTA ATAATTAACA TGTAATGCAAT GACGTTATT ATGAGATGGG TTTTATGAT TAGAGTCCC  
1121 CAATTAACA TTAAATAGC GATAAGAAAC AAAATATAGC GGCACAACTA GGATAAATTA TCGGGGGGG  
HindIII

HindIII

Clal

XbaI

1191 TGTCATCTAT GTTACTAGAT CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTGCC CTATAGTGA  
1261 TCGTATTACG CGCGCTCACT GGCGCTCGT TTACAACGTC GTGACTGGGA AACCCCTGGC GTTACCCAC  
1331 TTAATCCGCT TGAGGACAT CCCCTTTTCG CCAGCTGGCG TAATAGCGA GAGGCCGGCA CGGATCGCCC  
1401 TTCCCAACAG TTGCGCAGCC TGAAATGGCA ATGGGACGCG CCCTGTAGCG GGGCATTAAAG CGGGGGGGT  
1471 GTGGTGGTTA CGGGCAGCGT GACCGCTACA CTTGCCAGCG CCCTAGCGCC CGCTCCTTC GCCTTCTTCC  
1541 CTTCCCTTCT CGCCACGTC GCGGCTTTC CCCGTCAGC TCTAAATCGG GGGCTCCCT TAGGGTTCCG  
1611 ATTAGTGCT TTACGGCACC TGACCCCCA AAAACTTGAT TAGGGTGTAG GTTCACGTAG TGGGCCATCG  
1681 CCCTGATAGA CGGTTTTCG CCCCCTGAGC TTGGAGTCCA CGTTCTTTAA TAGGGACTC TTGTTCCAAA  
1751 CTGGAAACAA ACTCAACCT ATCTCGGTCT ATTCTTTGAA TTATAAGGG ATTTCGGCCTA TTTCGGCCTA  
1821 TTGGTTAAA ATGAGCTGA TTAAACAAA ATTTAACGG AATTAAACA AAATATAAC GCTTACAATT  
1891 TAGGTGGCAC TTTTCGGGA AATGTGCGCG GAACCCCTAT TTGTTTATT TTCTAAATAC ATCAAATAT  
1961 GTATCCGCTC ATGAGACAA AACCTGTCAA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT  
2031 CAACATTCC GTGTCGCCCT TATCCCTT TTTGGGCAT TTTGCCCTC TGTGTTGGCT CACCCAGAAA  
2101 CGCTGGTGAAG AGTAAAAGAT GCTGAAGAT AGTTGGGTGC ACGAGTGGGT TACATGGAAC TGGATCTCAA  
2171 CAGGGTAAAG ATCCTTGAGA GTTTCGGCCC CGAAGAACGT TTCCCAATGA TGAGGCACTT TAAAGTTCTG  
2241 CTATGGGG CGGTATTATC CGGTATTATC CGGAAACTCGG TCGCCGCAAG AGCAACTCGG CACTATTCTC  
2311 AGAATGACTT GTGTGAGTAC TCACCAAGTC CAGAAAAGCA TCTTACGGAT GGGCATGACAG TAAGAGAATT

FIG.-36C

2381 ATGGAGTGCT GCCATAACCA TGAGTGATAA CACTGGGCC AACTTACTTC TGACAACGAT CGGAGGACCG  
 2451 AAGGAGCTAA CGGCTTTTT GCACAACATG GGGGATCATG TAATCTGCCT TGATCGTGG GAAACGGAGC  
 2521 TGAATGAAAC CATAACCAAC GACGAGCGTG ACACCAAGAT GCCTGTAGCA ATGGCAACAA CGTTGGGCAA  
 2591 ACTATTAACCT GGGGAACTAAC TTACTCTAGC TTCCCGGCAA CAATTAAAGT ACTGGATGGA GGGGGATAAA  
 2661 GTTGCAAGGAC CACTTCTGGG CTCGGCCCTT CCGGCTGGCT GGTTTATTGC TGATAAATCT GGAGCCGGTG  
 2731 AGCGTGGTC TCGGGGTATC ATTGCAGCAC TGGTAAGGCC TCCCGTATCG TAGTTATCTA  
 2801 CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTG CTCACTGATT  
 2871 AAGCATGGT AACTGTCAGA CCAAGTTAAC TCATATATAC TTAGAGTTGA TTAAAAACTT CATTTTAAAT  
 2941 TTAAGGGAT CTAGGTGAAG ATCCTTTTG ATAATCTCAT GACCAAAATC CCTTAACGTG AGTTTTTCGTT  
 3011 CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT TCTTGAGATC CTTTTTTCT GCGCGTAATC  
 3081 TGCTGCTTGC AAACAAAAAA ACCACCGCTA CCAGGGGGTGG TTTGTTGCCC GGATCAAGAG CTACCAACTC  
 3151 TTTTCCGAA GGTAACTGGC TTCAAGGAGG CGCAGATAACC AAATAACTGTC CTTCTAGTGT AGCCGTAGTT  
 3221 AGGCCACCCAC TTCAAGAACT CTGTAGCAC GCCTACATAC CTCGCTCTGC TAATCCTGTT ACCAGTGGCT  
 3291 GCTGCCAGTG GCGATAAAGTC GTGTCTTACCG GGGTGGACT CAAGACGATA GTTACCGGAT AGGGCCAGC  
 3361 GGTGGCTG AACGGGGGT TCGTGCACAC AGCCAGCTT GGAGGAAACG ACCTACACCG AACTGAGATA  
 3431 CCTACAGCGT GAGCTATGAG AAAGGCCAC GCTTCCCAGA GGGAGAAGG CGGACAGGTA TCCGGTAAGC  
 3501 GGCAGGGTGC GAAAGGAGA GCGCACGAGG GAGCTTCCAG GGGAAACGC CTGGTATCTT TATACTCCTG  
 3571 TCGGGTTCTG CCACCTCTGA CTTGAGCGTC GATTTTTGTG ATGCTCGTCA GGGGGGGAA GCCTATGGAA  
 3641 AAACGCCAGC AACGGGGCCT TTATACGGTT CCTGGCCTT TGCTGCACAT GTCTTTCCCT  
 3711 GCGTTATCCC CTGATTCTGT GGATAAACGT ATTACCGCT TGAGTGAGC TGATAACGGCT CGCCGGAGCC  
 3781 GAAACGCCGA GCGCAGGGAG TCAGTGAGGG AGGAAGGGAA AGGGGCCCA ATACGAAAC CGGCTCTCCC  
 3851 CGCGCGTTGG CCGATTCAAT ATGGAGCTG GCACGACAGG TTTCCCGACT GAAAGGGGG CAGTGAGGCC  
 3921 AACGCAATTAA ATGTGAGTTA GCTCACTCAT TAGGCACCCC AGGCTTACA CTTTATGCTT CGGCTCGTA  
 3991 TGTGTTGAGG AATTGTGAGG GGATAACAAAT TTCAACACAGG AAACAGCTAT GACCATGATT AGGCCAAGCG

XbaI  
KpnI

4061 CGCAATAAC CCTCACTAA GGGAACAAA GCTGGGTACCG GGGCCCCCCC TCGAGGTCA TCATATGCTT  
 4131 GAGAAGGAG TCGGGATAGT CCAAAATAAA ACAAAAGTAA GATTACCTGG TCAAAGTGA AAACATCAGT  
 4201 TAAAGGTGG TATAAGTAAA ATATCGGTTA TAAAGGTGG CCCAAAGTGA ATTTCATCT TTCTCTACTAT  
 4271 TATAAAATT GAGGTGTTTG TGTCGGTACT TTGATGACGTC ATTTCATCT TAAAGTTGTT GAAATTGTT  
 4341 TCGCGATTG GAAATGCAAT TCTGTTATTG AGTCGGTTT TAAGTTGTT GCTTTTCGTT GCTTTTCGTT  
 4411 ATTGTGATAA GAAATATCTT TAAAGACATAAT TTGACATAAT TTTTGAGAAA AATATATATT

EcoRI

4481 CAGGGAAATT CCACAAATGAA CAATAATAAG ATTAAATAAG CTTGGCCCCCG TTGCAGGGAT GGGTATTTT

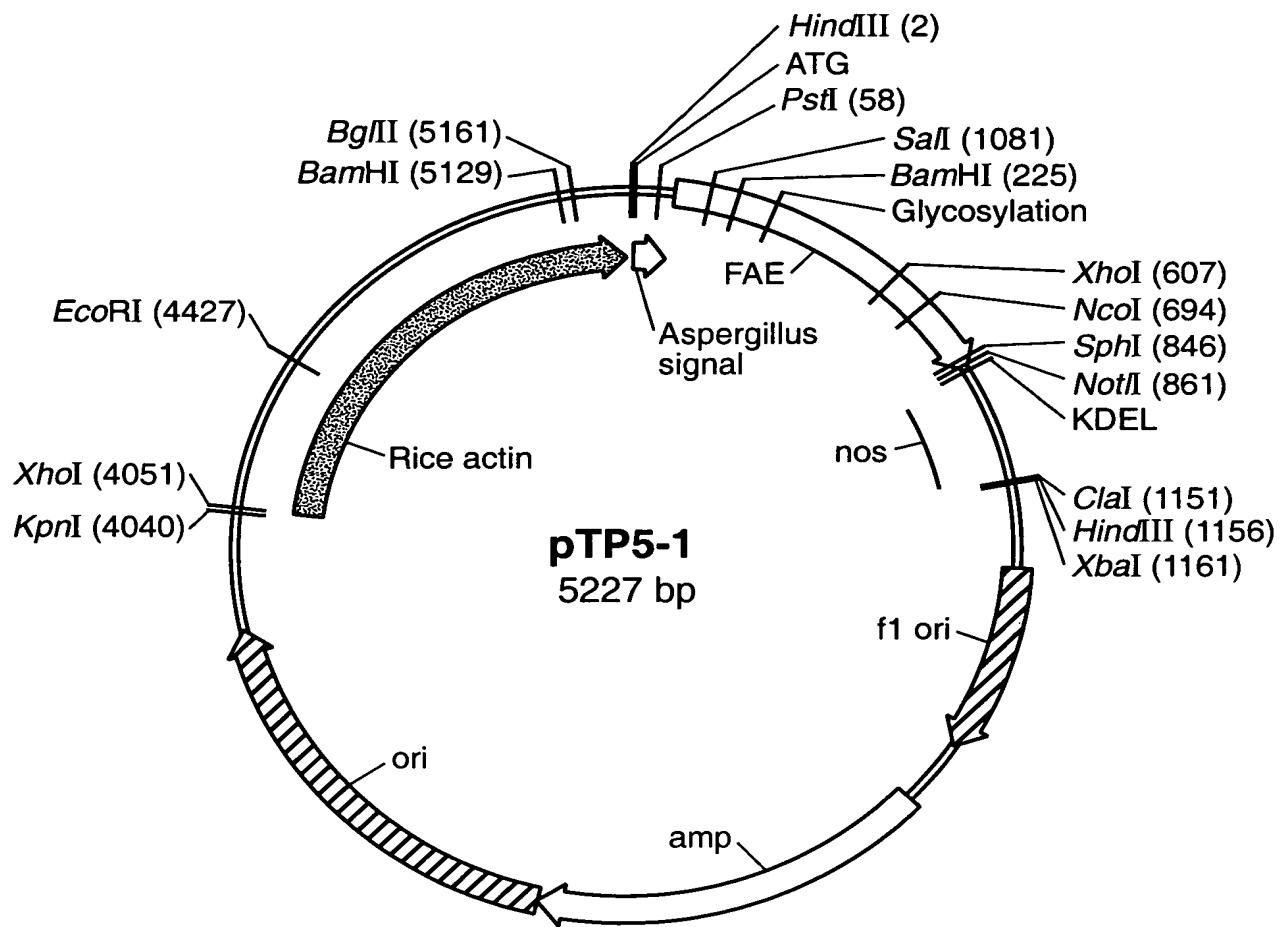
FIG.\_36D

4551 TCTAGTAAA TAAAGATAA ACTTAGACTC AAAACATTAA CAAAACAAC CCCTAAAGTC CTAAGCCCCA  
4621 AAGTGGCTATG CACGATCCAT AGCAAGCCCA GCCCCAACCCCA ACCCAACCCCA GTGCAGCCAA  
4691 CTGGCAAAATA GTCTCCACCC CCGGCACTAT CACCGTGAAT TGTCCGGCACC ACCGGCACTC TCGCAGCCAA  
4761 AAAAAAAAAGAAAGAAA AAAAGAAAAA AAAAGAAAAA AAAAGAAAAA AGAAACAGC AGGTGGTCCC GGTCGGAAAA  
4831 GCGAGGGAGGA TCGCGAGGAG CGACGAGGCC CGGGCCCTCCC TCCGCTTCCA AGAAACGCC CCCCATCGCC  
4901 ACTATAATACA TACCCCCCCC TCTTCTTCCC CCTAACCAAC CCTAACCAAC CACCTCCCTC  
4971 CCCCTCGCTG CCGGACGACG AGCTCCCTCCC CCCTCCCCCT CGGCAACCC ACCGAAACAC  
5041 TCCTCTTCT TCTCCGTTT TTTTTTTCGT CTCGGTCTCG ATCTTGGCC TTGGTACTTT GGTTGGCGA  
5111 GAGCGGGCTTC GTCGCCCCAGA TCGGTGGCGG GATCTCGGG CTTGGGTCTC CGGGCGTCTC  
BglII

BamHI

5181 TCGCCCCGGA TCCTCGGGG GAATGGGGCT CTCGGATGTA GATCTCTCTT CTTTCTTCTT TTTGTGGTAG  
5251 AATTGAAATC CCTCAGGATT GTTCATCGGT AGTTTTCTT TTCAATGATT GTGACAAATG CAGCCTCGTG  
5321 CGGAGCTTT TTGTAGC

**FIG.- 36E**



**FIG.\_37A**

HindIII

PstI

~~~~~  
 M K Q F S A K H V L A V V V T A G H A L A
 1 AAGCTTAACA TGAAGCAGTT CTCGCCAAA CACGTCCCTCG CAGTGCAGGG CACGCCCTAG
 . A S T Q G I S E D L Y S R L V E M A T I S Q A .
 71 CAGGCCTCTAC GCAAGGCATC TCCGAAGACCC TCTACAGCCG TTAGTGCAGA ATGGCCACTA TCTCCCAAGC
 Sall

~~~~~  
 A Y A D L C N I P S T I I K G E K I Y N S Q T  
 141 TGCCCTACGCC GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTACAA TTCTCAAATC  
 BamHI

~~~~~  
 D I N G W I L R D D S S K E I I T V F R G T G S
 GACATTAACG GATGGATCTT CCGCGACGAC AGCAGCAAAG AAATAATTCAC CGTCCTTCGCT GCCACACTGGTA
 . D T N L Q L D T N Y T L T P F D T L P Q C N G .
 281 GTGATAACGAA TCTACAACTC GATACTAACT ACACCCCTCAC GCCTTTGAC ACCCTTACAC AATGCAACGG
 . C E V H G G Y Y I G W V S V Q D Q V E S L V K
 351 TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AGTGCAGGTC GCTTGTCAAA
 Q Q V S Q Y P D Y A L T V T G H X L G A S L A A
 421 CAGCAGGTTA GCCAGTATCC GGACTACGGC CTGACCGTGA CGGGCAACRC CCTCGGGGCC TCCCTGGGG
 . I T A A Q L S A T Y D N I R L Y T F G E P R S .
 491 CACTCACTGC CGCCCAAGCTG TCTCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGGGCGAG

XbaI

~~~~~  
 G N Q A F A S Y M N D A F Q A S S P D T T Q Y  
 561 CGGCAATCAG GCCTTCGCGT CGTACATGAA CGATGCCCTTC CAAGGCCATCGA GCCCAGATAAC GACGCAGTAT  
 NcoI

~~~~~  
 F R V T H A N D G I P N L P P V E Q G Y A H G G
 631 TTCCGGGTCA CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCATGGCG
 . V E Y W S V D P Y S A Q N T F V C T G D E V Q .
 701 GTGTAGAGTA CTGGAGCGTT GATCCTTACA GGGCCAGAA CACATTGTCTGCACTGGGG ATGAAGGTGCA
 . C C E A Q G G Q G V N N A H T T Y F G M T S G
 771 GTGCTGTGAG GCCCAAGGGG GACAGGGGT GAATAATGCG CACACGACTT ATTGTTGGAT GACGAGGGCG

FIG.-37B

FIG._37C

| | | | | | | | |
|------|-------------|-------------|--------------|-------------|-------------|-------------|-------------|
| 2731 | TAGTTATCTA | CAAGCACGGGG | AGTCAGGCCAA | CTATGGATCA | ACGAAATAGA | CAGATCGCTG | AGATAAGGTGC |
| 2801 | CTCACTGATT | AAGCATTTGGT | AACTGTCAGA | CCAAGTTTAC | TCATATATAC | TTAGATTGA | TTAAAAACT |
| 2801 | CATTTTTAAT | TTAAAGGAT | CTAGGTGAAG | ATCCTTTTG | ATAATCTCAT | GACCAAATC | CCTTAACGTC |
| 2871 | AGTTTTCTGTT | CCACTGAGCG | TCAGACCCCG | TAGAAAAGAT | CAAAGGATCT | TCTGAGATC | CTTTTTTTCT |
| 2941 | GCGCGTAATC | TGCTGCTTGC | AAACAAAAAA | ACCACCGCTA | CCAGCGTTGG | TTGTTGTTGCC | GGATCAAGAG |
| 3011 | CTACCAAACTC | TTTTTCCGAA | GGTAACCTGGC | TTCAAGCAGAG | CCGAGATACC | AAATACTGTC | CTTCTAGTGT |
| 3081 | AGCCGTAGTT | AGGCCACCA | TTCAGAACT | CTGTAGCACC | GCCTACATAC | CTCGCTCTGC | TAATCCTGTT |
| 3151 | ACCAAGTGGCT | GCTGCCAGTG | GCGATAAGTC | GTGTCTTAC | GGGTTGGACT | CAAGACGATA | GTACCGGGAT |
| 3221 | AAGGGCGAGC | GGTGGGGCTG | AACGGGGGT | TGTTGCAACAC | AGCCCAGCTT | GGAGCGAACG | ACCTACACCG |
| 3291 | AACTGAGATA | CCTACAGCGT | GAGCTATGAG | AAAGGCCAC | GCTTCCCGAA | GGGAGAAAGG | CGGACAGGTA |
| 3361 | TCCGGTAAGC | GGCAAGGGTGC | GAACAGGAGA | GCGCAGGAGG | GAGCTTCCAG | GGGGAACGC | CTGGTATCTT |
| 3431 | TATAGTCCTG | TCGGGTTTCG | CCACCTCTGA | CTTGAGGTC | GATTTTTGTTG | ATGCTGTTCA | GGGGGGGGGA |
| 3501 | GCCTATGGAA | AAACGCCAGC | AACGCCGCCT | TTTACGGTT | CCTGGGCCTT | TGCTGGCCTT | TTGCTCACAT |
| 3571 | GTCTTTCT | GCATTATCCC | CTGATCTGT | GGATAACCGT | ATTACCGCT | TTGAGTGTAGC | TGATAACGCT |
| 3641 | CGCCGAGCC | GAACGACCGA | GCGCAGGGAG | TCAGTGAAGG | AGGAAGGGGA | ATACGCCAAC | AGAGGCCCA |
| 3711 | CGCCCTCTCCC | CGGCCGTTGG | CCGATTCATT | AATGCAGCTG | GCACGACAGG | TTTCCCGACT | GGAAAGGGGG |
| 3781 | CAGTGAGGC | AACGCAATT | ATGTGAGTTA | GCTCACTCAT | TAGGCACCCC | AGGCTTTACA | CTTTATGCTT |
| 3851 | CGGGCTCGTA | TGTTGTTGGC | AATTTGGTCA | GGATAACAAAT | TTCACACAGG | AAACAGCTAT | GACCATGATT |
| 3921 | | | | | | | XhoI |
| | | | | | | KpnI | |
| | | | | | | | |
| 3991 | ACGCCAAGCG | CGCAATTAAAC | CCTCACTAAA | GGGAACAAAA | GCTGGGTACCC | GGGCCCCCCC | TGAGGGTCA |
| 4061 | TCAATGCTT | GAGAAGAGAG | TGGGGATAGT | CCAAAATAAA | ACAAGGTTAA | GATACCTGG | TCAAAGTGA |
| 4131 | AAACATCAGT | TAAAGGTGG | TATAAGTAA | ATATCGTAA | TAAAAGTGG | CCCAAAGTGA | AATTACTCT |
| 4201 | TTCCTACTAT | TATAAAATT | GAGGATGTT | TGTGGTACT | TGATAACGTC | ATTTTGTAT | GAATGGTT |
| 4271 | TTAAGTTAT | TCGGGATTTCG | GAAATGCATA | TCTGTTTTG | AGTCGGTTT | TAAGTTGCTT | GCTTTGTAA |
| 4341 | ATACAGAGGG | ATTGTATAA | GAATATCTT | TAAAAACCC | ATATGCTAA | TTGACATAAT | TTTTGAGAAA |
| | | | | | | | EcoRI |
| | | | | | | | |
| 4411 | AATATATATT | CAGGGAAATT | CCACAAATTGAA | CAATAATAAG | ATTAAATAG | CTTGGCCCCG | TGGCAAGCGAT |
| 4481 | GGGTATTTT | TCTAGTAAA | TAAAGATAA | ACTTAGACTC | AAAACATTAA | CAAAACAAAC | CCCTAAAGTC |
| 4551 | CTAAAGGCCA | AAGTGCTATG | CACGATCCAT | AGCAAGCCCA | GCCCAACCCAA | ACCCAACCCAA | ACCCACCCCA |
| 4621 | GTGCAGGCCAA | CTGCCAAATA | GTCTCCACCC | CGGGCACTAT | CACCGTGAAGT | TGTCGGCACCC | ACCGCACGTC |
| 4691 | TCGGAGCCAA | AAAAAAATAAA | AGAAAGAAAAA | AAAAAACAGC | AGGTGGGTTC | GGGTGCTGGCC | GGGTGCTGGCC |

FIG. 37D

4761 GGCCGGAAAA GCGAGGAGGA TCGGGAGGAG CGACGAGGCC CGGGCCTCCC TCGGCTTCGA AAGAAACGCC
4831 CCCCATCGCC ACTATATACA TACCCCCCCC TCTCCCTCCA CCTTACCAAC CCACCCACCAC
4901 CACCTCCCTCC CCCCCTCGCTG CGGACGACG AGCTCTCCC CCCTCCCCCT CGGCCGCCGC CGGTAACCAC
4971 CCCGCCCTC TCCTCTTTCT TTCTCCGTTT TTTTTCTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT
5041 GGGTGGCGGA GAGCGGCTTC GTCCGCCAGA TCGGCCAGA GGAGGGGGGG GATCTCGGGG CTGGCGTCTC
BamHI ~~~~~
5111 CGGGCGTGTAG TCGGGCCGGA TCCTCGGGG GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT
5181 TTGTTGGTAG AATTGAAATC CCTCAGCATT GTTCATCGGT AGTTTTCTT TTCAATGATT GTGACAAATG
5251 CAGCCTCGTG CGGAGCTTT TTGTAGC

FIG._37E

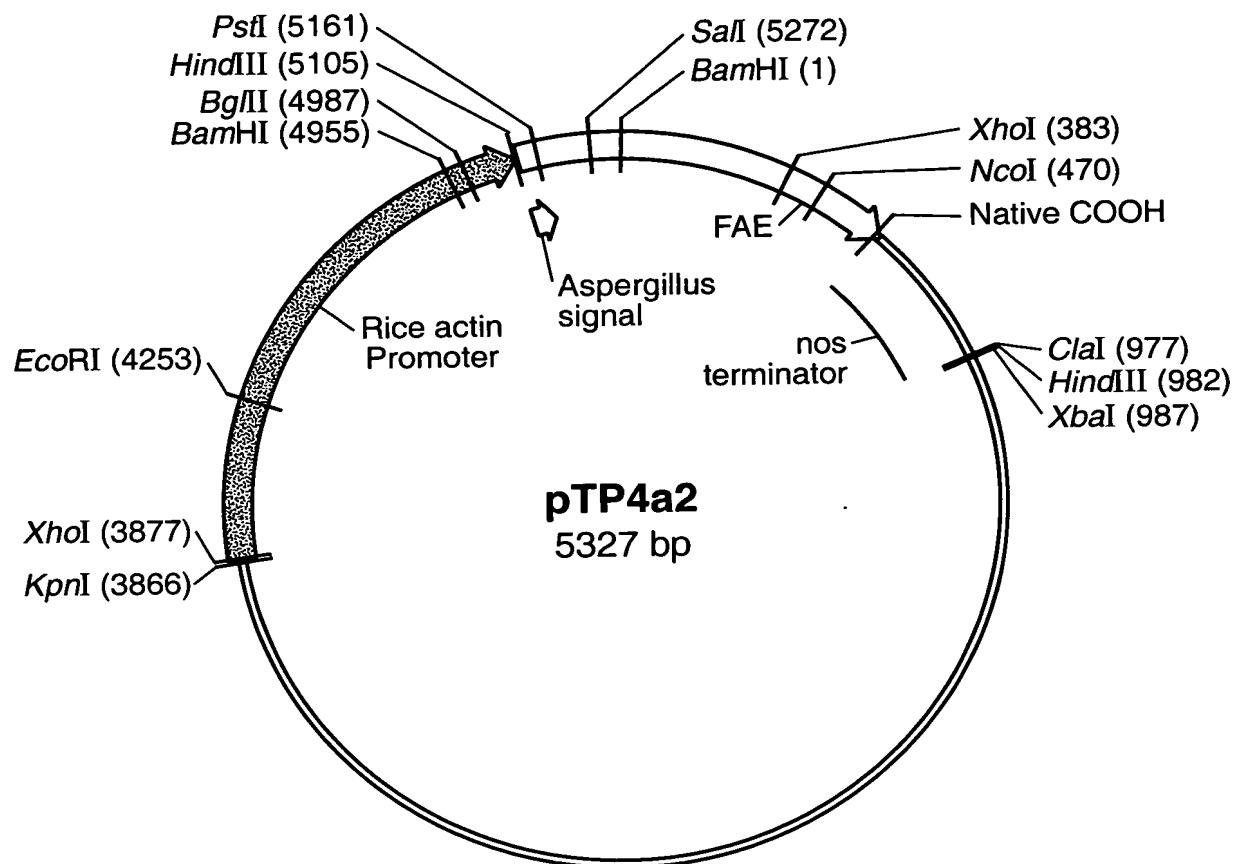


FIG._38A

BamHI

• I L R D D S S K E I I T V F R G T G S D T N L
1 GATCCTCCGC GACGACAGCA GCAAAGAAAT AATCACCGTC TTCCGGTGGCA CTGGTAGTGA TACGAATCTA
Q L D T N Y T L T P F D T L P Q C N G C E V H G
71 CAACTCGATA CTTAACTACAC CCTCACGGCCT TTCGACACCC TACCAAAATG CAAACGGTTGT GAAGTACACG
. G Y I G W V S V Q D Q V E S L V K Q Q V S Q .
141 GTGGATATTA TATTGGATGG GTCTCCGGTCC AGGACCAAGT CGAGTCGGCTT GTCAAACAGC AGGGTTAGGCCA
. Y P D Y A L T V T G H X L G A S L A A L T A A
211 GTATCCGGAC TAGGGCTGA CCGTGACCGG CCACKCCCTC GGCGCCTCCC TGGGGCACT CACTGCCGCC
Q L S A T Y D N I R L Y T F G E P R S G N Q A F
281 CAGCTGTCTG CGACATACGA CAACATCCGC CTGTACACCT TCGGCGAACCGC TCGGCGAACCC AATCAGGGCCT
XbaI

• A S Y M N D A F Q A S S P D T T Q Y F R V T H .
351 TCGCGTCGTA CATGAACGAT GCCCTTCCAAG CCTCGAGGCC AGATACGACG CAGTATTTC GGGTCACTCA
NcoI

• A N D G I P N L P P V E Q G Y A H G G V E Y W
421 TGCCAACGAC GGATCCCAA ACCTGCCCAAC GGTGGAGCAG GGGTACGCCCGGGTACGCCCTT AGAGTACTGG
S V D P Y S A Q N T F V C T G D E V Q C C E A Q
491 AGCGTTGATC CTTACAGGGC CCAGAACACA TTTGTCTGCA ATCGATGTTGAA TTACATGTA
. G G Q G V N N A H T T Y F G M T S G A C T W * .
561 AGGGGGGACA GGGTGTGAAT AATGGCCACA CGACTTATT TGGGATGACG AGGGGAGCCT GTACATGGTG
* *

631 ATCAGTCATT TCAGCCTCCC CGAGTGTACG AGGAAAGATG GATGTCCTGG AGAGGGGGCC GCGTAACCAAC
701 TGAAGGATGA GCTGTAAGA AGCAGATCGT TCAAACATT GGCAATAAAG TTCTCTTAAGA TTGAAATCCTG
771 TTGCGGGTCT TGCGATGATT ATCATATAAT TTCTGTGAA TTACATGTA CATGTAATAA TTACATGTA
841 ATGCATGACG TTATTTATGA GATGGGTTT TATGATTAGA GTCCCGAAAT TATACATTTA ATACGGGATA
911 GAAAACAAAA TATAGGGCCG AACTAGGAT AAATTATCGC GGGGGGTGTC ATCTATGTTA CTAGATCGAT
XbaI

~~~~~  
HindIII  
~~~~~

FIG.-38B

2009-07-02 09:12:09

67 / 154

| | | | | | | | | | | | |
|------|-------------|-------------|-------------|-------------|-------------|--------------|-------------|-------------|-------------|-------------|-------------|
| 981 | AAGCTTCTAG | GGGGCCGGT | GGAGCTCCAA | TTCGCCCCAA | TTAGTGGTGT | ATTAGGCTTAT | AGTGGCTCTAT | CTCACTGGCG | CTTAGCGCG | CTTAGCGCG | CTCACTGGCC |
| 1051 | GTCGTTTAC | AACGTCGTGA | CTGGGAAAAC | CCCAACTTAA | TGCGCGTTA | CCCAACTTAA | TGCGCGTTA | GCACATCCCC | GCACATCCCC | GCACATCCCC | GCACATCCCC |
| 1121 | CTTTCGCCAG | CTGGCGTAAT | AGCGAAGAGG | CCCGCACCGA | TGCGCCTTCC | CAACAGTTC | GCAGGCCTGAA | GCAGGCCTGAA | GCAGGCCTGAA | GCAGGCCTGAA | GCAGGCCTGAA |
| 1191 | TGGCGAATGG | GACGCCCT | GTAGCGGCCG | ATTAAGCGCG | GGGGGTGTGG | TGGTTACGGG | CAGGGTGAAC | CAGGGTGAAC | CAGGGTGAAC | CAGGGTGAAC | CAGGGTGAAC |
| 1261 | GCTACACTTG | CCAGGCCCT | AGGCCCGCT | CCTTTCGCTT | TCTTCCCTTC | CTTCTCGCC | ACGTTCGCCG | ACGTTCGCCG | ACGTTCGCCG | ACGTTCGCCG | ACGTTCGCCG |
| 1331 | GCTTCCCCG | TCAAGCTTA | AATGGGGGC | TCCCTTGG | GTTCCGATT | AGTGCCTTAC | GGCACCTCGA | GGCACCTCGA | GGCACCTCGA | GGCACCTCGA | GGCACCTCGA |
| 1401 | CCCCAAGAA | CTTGATTAGG | GTGATGGTT | ACGTAGTGGG | CCATCGCCCT | GATAGACGGT | TTTTCGCCCT | TTTTCGCCCT | TTTTCGCCCT | TTTTCGCCCT | TTTTCGCCCT |
| 1471 | TTGACGTTGG | AGTCCACGTT | CTTAATAGT | GGACTCTTGT | TCCAACCTGG | AACAACACTC | AACCCCTATCT | AACCCCTATCT | AACCCCTATCT | AACCCCTATCT | AACCCCTATCT |
| 1541 | CGGTCTATT | TTTGATTAA | TAAGGGATT | TGCCGATTT | GGCCTATTTG | TAAAAAAATG | AGCTGATTAA | AGCTGATTAA | AGCTGATTAA | AGCTGATTAA | AGCTGATTAA |
| 1611 | ACAAAATT | AACGCGAATT | TTAACACAAAT | ATTAACGCTT | ACAATTAGG | TGGCACTTT | CGGGAAATG | CGGGAAATG | CGGGAAATG | CGGGAAATG | CGGGAAATG |
| 1681 | TGGCGGAAC | CCCTATTGT | TTATTTTCT | AAATACATT | AAATATGTAT | CCGTCATGA | GACAAATAACC | GACAAATAACC | GACAAATAACC | GACAAATAACC | GACAAATAACC |
| 1751 | CTGATAAATG | CTTCAATAAT | ATTGAAAAG | GAAGAGTATG | AGTATTCAAC | ATTTCGTTGT | CGCCCTTATT | CGCCCTTATT | CGCCCTTATT | CGCCCTTATT | CGCCCTTATT |
| 1821 | CCCTTTTG | GGGCATTG | CCTTCCTGTT | TTTGCCTACC | CAGAAACGCT | GGTGAAGGTA | AAAGATGCTG | AAAGATGCTG | AAAGATGCTG | AAAGATGCTG | AAAGATGCTG |
| 1891 | AAGATCAGTT | GGGTGCAAGA | GTGGGTTACA | TCGAACCTGGA | TCTCAACAGC | GGTAAGATTC | TGAGAGAGTT | TGAGAGAGTT | TGAGAGAGTT | TGAGAGAGTT | TGAGAGAGTT |
| 1961 | TGGCCCCGAA | GAACGTTTC | CAATGATGAG | CACTTTAAA | GTTCCTGCTAT | GTGGCGGGT | ATTATCCGT | GTGGCGGGT | ATTATCCGT | GTGGCGGGT | ATTATCCGT |
| 2031 | ATTGACGCCG | GGCAAGAGCA | ACTGGTGC | CGCATACACT | ATTCTCAGAA | TGACTTGGT | GAGTACTCAC | GAGTACTCAC | GAGTACTCAC | GAGTACTCAC | GAGTACTCAC |
| 2101 | CAGTCACAGA | AAAGCATCTT | ACGGATGGCA | TGACAGTAAG | AGAATTATGC | AGTGTGC | TAACCATGAG | TAACCATGAG | TAACCATGAG | TAACCATGAG | TAACCATGAG |
| 2171 | TGATAACACT | GGGGCAACT | TACCTTCTGAC | AACGATCGGA | GGACCGAAGG | AGCTAACCGC | TTTTTGAC | TTTTTGAC | TTTTTGAC | TTTTTGAC | TTTTTGAC |
| 2241 | AACATGGGG | ATCATGTAAC | TCGCGCTTGAT | CGTGGGAAC | CGGAGCTGAA | TGAAGCCATA | CCAACAGACG | CCAACAGACG | CCAACAGACG | CCAACAGACG | CCAACAGACG |
| 2311 | AGCGTGACAC | CACGATGCT | GTAGCAATGG | CAACAACGTT | GGCAAACTA | TAACTGGCG | AACTAATTCAC | AACTAATTCAC | AACTAATTCAC | AACTAATTCAC | AACTAATTCAC |
| 2381 | TCTAGCTTCC | CGGCAACAAAT | TAATAGACTG | GATGGAGGGC | GATAAAGTTG | CAGGACCAT | TCTGCGCTCG | TCTGCGCTCG | TCTGCGCTCG | TCTGCGCTCG | TCTGCGCTCG |
| 2451 | GCCCTTCGG | CTGGCTGGTT | TATGCTGAT | AAATCTGGAG | CGGGTGAGCG | TGGGTCTCG | GGTATCATTTG | GGTATCATTTG | GGTATCATTTG | GGTATCATTTG | GGTATCATTTG |
| 2521 | CAGGACTGGG | GGCAGATGGT | AAGCCCTCCC | GTATCCTGAT | TATCTACAGC | ACGGGGAGTC | AGGCAACTAT | AGGCAACTAT | AGGCAACTAT | AGGCAACTAT | AGGCAACTAT |
| 2591 | GGATGAACGA | ATAGACAGA | TCGCTGAGAT | AGGTGCCTCA | CTGATTAAAC | ATGGTAAC | GTCAAGACCA | GTCAAGACCA | GTCAAGACCA | GTCAAGACCA | GTCAAGACCA |
| 2661 | GTTTACTCAT | ATATACTTTA | GATGGATTAA | AAACTCTATT | TTAAATTAA | AAGGATCTAG | GTGAAGATTC | GTGAAGATTC | GTGAAGATTC | GTGAAGATTC | GTGAAGATTC |
| 2731 | TTTTTGATAAA | TCTCATGACC | AAAATCCCTT | AACGTTGAGTT | TTCGTTCCAC | TGAGCGTGTAG | ACCCCGTAGA | ACCCCGTAGA | ACCCCGTAGA | ACCCCGTAGA | ACCCCGTAGA |
| 2801 | AAAGATCAAA | GGATCTTCTT | GAGATCCCTT | TTTTCTGCGC | GTAAATCTGCT | GCTTGGCAAAAC | AAAAAAACCA | AAAAAAACCA | AAAAAAACCA | AAAAAAACCA | AAAAAAACCA |
| 2871 | CCGCTACCAAG | CGGTGGTTG | TTTGGCGGAT | CAAGAGCTAC | CAACTCTTT | TCCGAAGGTA | ACTGGCTTCA | ACTGGCTTCA | ACTGGCTTCA | ACTGGCTTCA | ACTGGCTTCA |
| 2941 | GCAGAGGGCA | GATAACAAAT | ACTGTCCCTTC | TAGTGTAGCC | GTAGTTAGGC | CACCACTCA | AGAACTCTGT | AGAACTCTGT | AGAACTCTGT | AGAACTCTGT | AGAACTCTGT |
| 3011 | AGCACCGCCT | ACATACCTCG | CTCTGCTTAAT | CCTGTTACCA | GTGGCTGCTG | CCAGTGGCGA | TAAGTCGTGT | TAAGTCGTGT | TAAGTCGTGT | TAAGTCGTGT | TAAGTCGTGT |
| 3081 | CTTACCGGGT | TGGACTCAAG | ACGATAGTTA | CCGGATAAGG | CCGAGGGTC | GGGCTGAACG | GGGGTTTCGT | GGGGTTTCGT | GGGGTTTCGT | GGGGTTTCGT | GGGGTTTCGT |
| 3151 | GCACACAGCC | CAGCTGGAG | CGAACGACCT | ACACCGAACT | GAGATAACCA | CAGCGTGAAC | TATGAGAAG | TATGAGAAG | TATGAGAAG | TATGAGAAG | TATGAGAAG |
| 3221 | CGCCACGGCTT | CCCGAAGGGA | GAAGGGGA | CAGGTATCCG | GTAAGGGGGCA | GGGTGGGAAC | AGGAGAGGCG | AGGAGAGGCG | AGGAGAGGCG | AGGAGAGGCG | AGGAGAGGCG |
| 3291 | ACGAGGGAGC | TTCCAGGGGG | AAACGGCTGG | TATCTTATA | GTCCCTGTCGG | CTCTGACTTGT | CTCTGACTTGT | CTCTGACTTGT | CTCTGACTTGT | CTCTGACTTGT | CTCTGACTTGT |
| 3361 | AGCGTCCGATT | TTTGTGATGC | TCGTCAAGGG | GGGGAGGCCT | ATGGAAAAC | GCCAGCAACG | CGGGCTTTT | CGGGCTTTT | CGGGCTTTT | CGGGCTTTT | CGGGCTTTT |

FIG.-38C

3431 ACGGTTCTG GCCTTTGCT GGCCTTTGCT TCACATGTC TATCCCTCGT TATCCCTGCA TATCCCTCGT TATCCCTGCA
3501 AACCGTATTA CGCCCTTGA GTGAGCTGAT ACCGCTCGCC GACGCCGAAAC AGCGAGTCAG
3571 TGAGCGAGGA AGCGGAAGAG CGCCCAATAC GCAAACCGCC TCTCCCGCG CCGTGGCCGA TTCTATTAAATG
3641 CAGCTGGCAC GACAGGTTTC CCGACTGGAA AGCGGGCACT GAGCGCAACG CAAATTAGT GAGTTAGCTC
3711 ACTCATAGG CACCCAGGC TTTACACTT ATGCTTCGG CTCGTTATGTT GTGTGGAAAT GTGAGGGAT
3781 AACAAATTCA CACAGGAAAC AGCATGACC ATGATTACGC CAAGCGGCA ATTAACCTC ACTAAAGGGA

KpnI

3851 ACAAAAGCTG GGTACCGGGC CCCCCCTCGA GGTCAATTCA ATGCTTGA AGAGAGTCGG GATAGTCCAA
3921 AATAAAACAA AGGTAAAGATT ACCTGGCTAA AAGTGAACAC ATCAGTAAA AGGTGGTATA AGTAAATAT
3991 CGGTAATAAA AGGTGGCCA AAGTGAATT TACTCTTTC TACTATTATA AAAATTGAGG ATGTTTTGTC
4061 GGTACTTTGA TAGTCAATT TTGTATGAAT TGGTTTTAA GTTATTCGC GATTGGAAA TGCATATCTG
4131 TATTTGAGTC GGTTTTAAG TTCTGGCTT TTGTAAATAC AGAGGGATT GTATAAGAAA TATCTTTAAA

EcoRI

4201 AAACCCATAT GCTAATTGTA CATAATTGTA GAGAAAATA TATATTCAAG CGAATTCCAC ATGAAACAT
4271 AATAAGATTA AAATAGCTG CCCCGTTGC AGCGATGGT ATTTTCTA GTAAAATAAA AGATAAAACTT
4341 AGACTCAAAA CATTACAA AACAAACCT AAAGTCTAA AGCCCAAAGT GCTATGCAGG ATCCATAGCA
4411 AGCCCAAGCCC AACCCAAACCC ACCCCAGTGC AGCCAACTGG CAAATAGTCT CCACCCCCGG
4481 CACTATACC GTGAGTTGTC CGCACCAACG CACGTCCTCGC AGCCAAAAA AAAAAAGAA AGAAAAGAA
4551 GAAAAAGAAA AACAGGAGGT GGGTCCGGGT CGTGGGGGCC GGAAAAGCGCA GAGCAGCGAC
4621 GAGGCCGGC CCTCCCTCG CTTCACAAAGA AACGCCCTT ATCGGCCACTA TATACATACC CCCCCCTCTC
4691 CCTCCATCCC CCAAAACCTA CCACCAACAC CACCAACC ACCACCCAC CTCGCTGCCGG ACGACGAGCT
4761 CCTCCCCCTT CCCCCCTCCG ACCACCCCG CGCCGCCGGT TACGTTGGGT CCCCTCTCCCT CTTTCTTCTC
4831 TTCTGCTCG GTCTCGATCT TTGGCCTTGG TAGTTGGGT GGGCAGAGC GCTTCGTCG CCCAGATCGG
BamHI

4901 TGGCGGGAG GGGGGGATC TCGGGGCTG CGTCTCCGGG CGTGAGTCGG CCCGGATCCT CGGGGGAAAT
BglII

4971 GGGCTCTCG GATGTAGATC TTCTTTCTT CTTCCTTTTG TGTTAGAATT TGAATCCCTC AGCATTGTTCTC
HindIII

5041 ATCGGTAGTT TTCTTTTCA TGATTGTGA CAAATGGCAGC CTCGTCGGGA GCTTTTTGT AGCAAGCTTA

FIG._38D

69 / 154

5111 M K Q F S A K H V L A V V V T A G H A L A A S .
ACATGAAGCA GTTCTCGGCC AACACGTCC TCGCAAGTGT GGTGACTGCA GGGCACGCTC TAGCAGCCTC
· T Q G I S E D L Y S R L V E M A T I S Q A A Y
5181 TACGCAAGGC ATCTCCGAAG ACCTCTACAG CCGTTAGTC GAAATGGCCA CTATCTCCCA AGCTGGCTAC
SallI
~~~~~

5251 A D L C N I P S T I I K G E K I Y N S Q T D I N  
GCCGACCTGT GCAACATTCC GTGCACTATT ATCAAGGGAG AGAAAATTAA CAATTCTCAA ACTGACATTA  
B  
~~~~~

5321 · G W
ACGGATG

FIG.-38E

70 / 154

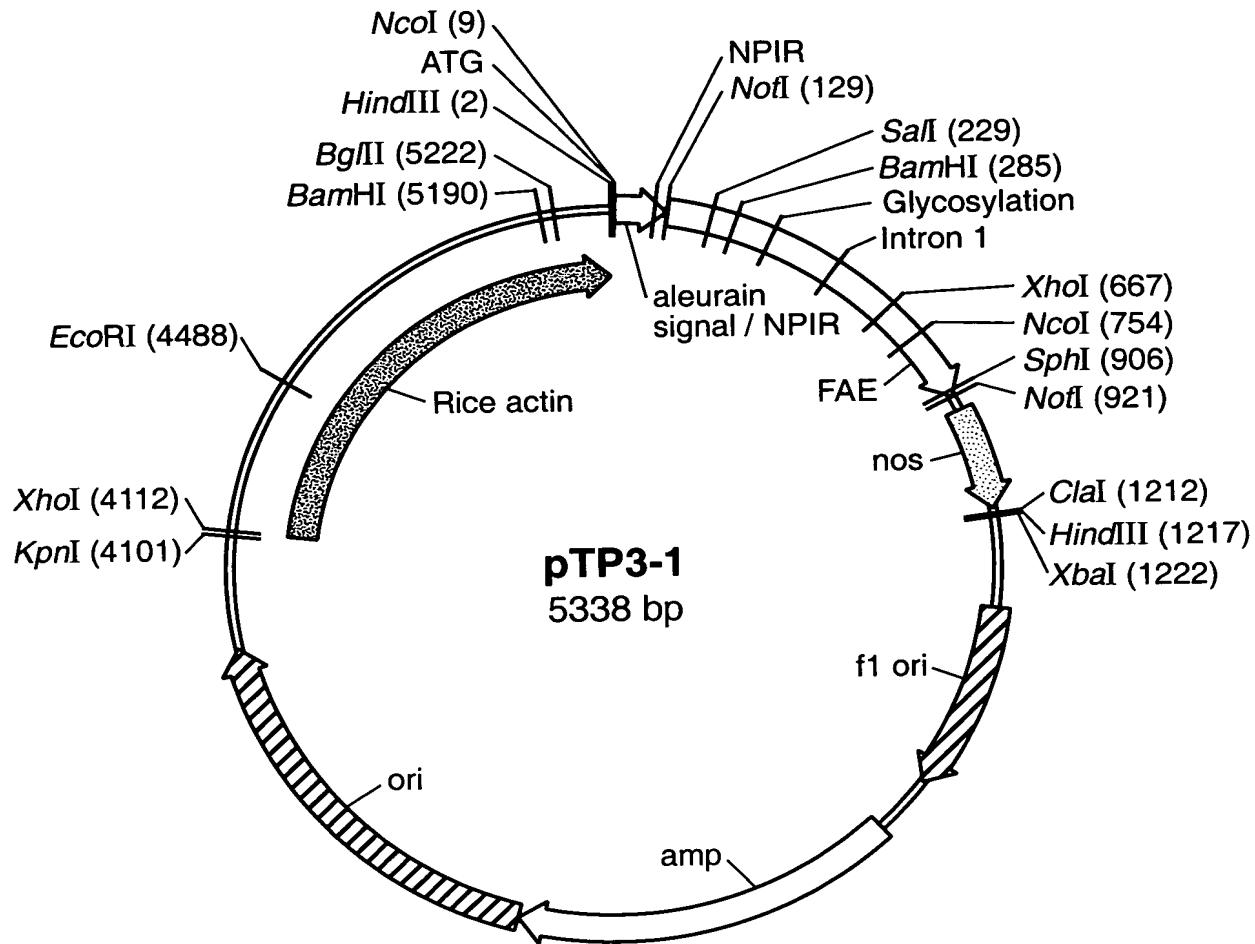


FIG._39A

NcoI

HindIII

1 AAGCTTACCA M A H A R V L L L A L A V L A T A A V A V
GGCCCTCCACGC CGCGGTCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCGTGGCG

NPIR

1 A S S S S F A D S N P I R P V T D R A A S T .
71 TCGCCTCCTC CTCCTCCTC GCCGACTCCA ACCCGATCCG GCCCCTACCC GACCGCGGG CGGCCTCCAC
. Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAAGACC TCTACAGGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCTTACGCC

Sall

211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATAATCAC CGTCTCCGT GGCACTGGTA GTGATAACG
. W I L R D D S S K E I I T V F R G T G S D T N .
281 GATGGATCCT CGCGACGAC AGCAGCAAAG AAATAATCAC TTCTCAAAC T GACATTAACG

BamHI

351 TCTACAACTC GATACTAATC ACACCCCTCAC GCCTTTGAC ACCCTACCCAC ATGCAACGG TTGTGAAGTA
H G G Y Y I G W V S V D Q V E S L V K Q Q V S
421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA
. Q Y P D Y A L T V T G H X L G A S L A A L T A .
491 GCCAGTATCC GGACTACGG CTGACCGTGA CGGGCACKC CCTCGGGCC TCCCTGGGG CACTCACTGC
. A Q L S A T Y D N I R L Y T F G E P R S G N Q
561 CGCCCAAGCTG TCTGCGACAT ACGACAAACAT CCGCCCTGTAC ACCTTCGGCG AACCCGCGAG CGGCAATCAG

Glycosylation

FIG._39B

XbaI

631 A F A S Y M N D A F Q A S S P D T T Q Y F R V T
GCCTTCGGGT CGTACATGAA CGATGCCCTTC CAAGCCTCGA GCCCAGATA GACGCAGTAT TTCCGGGTCA

NcoI

701 H A N D G I P N L P P V E Q G Y A H G V E Y .
CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCCATGGG GTGGTAGAGTA
• W S V D P Y S A Q N T F V C T G D E V Q C C E
771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTTGTC TGCACTGGG ATGAAAGTGA GTGCTGTGAG

SphI

841 A Q G G Q G V N N A H T T Y F G M T S G A C T W
GCCAGGGCG GACAGGGTGT GAATAATGCC CACACGACTT ATTTGGGAT GACGAGGGC GCATGCGACCT

NotI

911 P V A A E T T E G *
GGCCGGTCGC GGCGCGGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAAA CATTGGCAA
981 TAAAGTTCT TAGATTGAA TCCGTGCGA GGTCTTGCAT TGATTATCAT ATAATTCTG TTGAAATTACG
1051 TTAAGCATGT ATTAATTAC ATGTAATGCC TGACGTTATT TATGAGATGG GTTTTTATGA TTAGAGTCCC
1121 GCAATTATAAC ATTAATAAC CGATAGAAA CAAAATATAG CGCGCAACT AGGATAATT ATCGCGCGCG

HindIII

XbaI

1191 GTGTCATCTA TGTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCAATTCCG CCTATAGTGA
1261 GTCGTATTAC GCGCGCTCAC TGGCCGCTCGT TTTACAAACGT CGTGACTGGG AAAACCTGG CGTTACCCAA
1331 CTTAATGCC TTGCAAGCACA TCCCCCTTC GCCAGCTGGC GTAATAGCGA AGAGGCCGC ACGATCGCC
1401 CTTCCCAACA GTTGGCAGC CTTGAATGGCG AATGGGACGC GCCCTGTAGC GCGGCATAA GCGGGGGG

FIG.-39C

| | | | | | | | |
|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1471 | TGTGGTGGTT | ACGGCGAGGG | TGACCGGCTAC | ACTGGCCAGC | GCCCTAGCGC | CCGCTCCCTT | CGCTTCTTC |
| 1541 | CCTTCCTTTC | TGCCAACGTT | CGCCGGCTT | CCCCGTCAG | CTCTAAATCG | GGGGCTCCCT | TAGGGTTTC |
| 1611 | GATTTAGTGC | TTAACGGCAC | CTCGACCCCA | AAAACCTTGA | TTAGGGTGAT | GGTTCACGTA | GTGGGCCATC |
| 1681 | GCCCTGTAG | ACGGTTTTC | GCCCTTGTAC | GGTGGAGTCC | ACGTTCTTTA | ATAGTGGACT | CTTGTGTC |
| 1751 | ACTGGAACAA | CACTCAACCC | TATCTCGGTC | TATTCTTTG | ATTATAAGG | GATTTGCGG | ATTCTGGCC |
| 1821 | ATTGGTTAAA | AAATGAGGTG | ATTAAACAA | AATTAAACG | GAATTAAAC | AAAATATAA | CGCTTACAAT |
| 1891 | TTAGGGGCCA | CTTTTGGGG | AAATGTGCGC | GGAAACCCCTA | TTGGTTATT | TTCTCTAAATA | CATCCAATA |
| 1961 | TGTATCCGCT | CATGAGACAA | TAATGAGCAA | AAATCCCTGAT | AAATGTTCA | ATAATATTGA | AAAAGGAAGA |
| 2031 | TCAACATTTTC | CGTGTGCCC | TTATTCCTT | TTTGCGGCA | TTTGCTTTC | CTGTTTTG | TCACCCAGAA |
| 2101 | ACGCTGGTGA | AAGTAAAAGA | TGCTGAAGAT | CAGTGGGTG | CACGAGTGGG | TTACATCGAA | CTGGATCTCA |
| 2171 | ACAGGGTAA | GATCCTTGAG | AGTTTTCGCC | CGGAGAACG | TTTTCCAATG | ATGAGCACTT | TTAAAGTTCT |
| 2241 | GCTATGGGC | GCGGTATTAT | CCCGTATTGA | CGCCGGCAA | GAGCAACTCG | GTGCGCGCAT | ACACTATTCT |
| 2311 | CAGAATGACT | TGGTTGAGTA | CTCACCGTC | ACAGAAAAGC | ATCTTACGGA | TGGCATGACA | GTAAAGAGAA |
| 2381 | TATGGAGTGC | TGCCATAACC | ATGAGTGTAA | ACACTGGGC | CAACTTACTT | CTGACAACGA | TGGGAGGAC |
| 2451 | GAAGGAGCTA | ACCGCTTTT | TGACAACAT | GGGGATCAT | GTAACCTGCC | TTGATCGTTG | GGAAACGGAG |
| 2521 | CTGAATGAG | CCATACAAA | CGACGAGCGT | GACACCACGA | TGCCTGTAGC | ATGGCAACA | ACGTTGGCGCA |
| 2591 | AACTTTAAC | TGGGAACTA | CTTAACCTAG | CTTCCCGGCA | ACAATTAAATA | GAETGGATGG | AGGCGGATAAA |
| 2661 | AGTTGAGGA | CCACTTCTGC | GCTGGGCCCT | TCCGGCTGGC | TGTTTATTG | CTGATAAATC | TGGAGCCGGT |
| 2731 | GAGCGTGGGT | CTCGGGTAT | CATTCAGCA | CTGGGGCCAG | ATGGTAAGCC | CTCCCGTATC | GTAGTTATCT |
| 2801 | ACACGACGGG | GAGTCAGGCA | ACTATGGATG | AACGAAATAG | ACAGATCGCT | GAGATAGGTG | CCTCACTGAT |
| 2871 | TAAGCATTTG | TAACTGTAG | ACCAAGTTA | CTCATATAAA | CTTTAGATTG | ATTAAAATC | TCATTTTAA |
| 2941 | TTTAAAGGA | TCTAGGTGAA | GATCCTTTT | GATAATCTCA | TGACCAAAT | CCCTTAACGT | GAGTTTTCGGT |
| 3011 | TCCACTGAGC | GTCAAGACCC | GTAGAAAAAG | TCAAAAGGATC | TTCTTGAGAT | CCTTTTTTC | TGCGCGTAAT |
| 3081 | CTGCTGCTTG | CAAACAAAAA | AACCAACGGCT | ACCAGGGTG | GTTCGTTGC | GGGATCAAGA | GCTACCAACT |
| 3151 | CTTTTTCGGA | AGGTAACCTG | CTTCAGCAGA | GGCGAGATAAC | CAAATACGT | CCTCTCTAGT | TAGCCGTAGT |
| 3221 | TAGGCCACCA | CTTCAGAAC | TCTGTAGCAC | CGCCTACATA | CCTCGCTCTG | CTAACCCCTG | TACCACTGTC |
| 3291 | TGCTGCCAGT | GGCGATAAGT | CGTGTCTTAC | CGGGTTGGAC | TCAAGACGAT | AGTTACGGGA | TAAGGGCAG |
| 3361 | CGGTGGGCT | GAACGGGGGG | TTCGTGCACA | CAGCCCCAGCT | TGGAGCGAAC | GAACCTACACC | GAACGTGAGAT |
| 3431 | ACCTACAGCG | TGAGCTATGA | GAAAGCGCCA | CGCTTCCCGA | AGGGAGAAAG | GGGGACAGGT | ATCCGGTAAG |
| 3501 | CGGCAGGGTC | GGAACAGGGAG | AGGCCACGAG | GGAGCTTCCA | GGGGAAACG | CCTGGTATCT | TTATAAGTCCT |
| 3571 | GTCGGGTTTC | GCCACCTCTG | ACTTGGCGT | CGATTTTGT | GATGCTCGTC | AGGGGGGG | AGCCTATGGAA |
| 3641 | AAAACGCCAG | CAACGGGGCC | TTTTACGGT | TCCTGGCCTT | TGCTGGCCTT | TTTGCTTTC | TGTTGCTCACA |
| 3711 | TGCCTTATCC | CCTGATTCTG | TGGATAACCG | TATTACGCC | TTTGAGTGG | CTGATAACGCC | TGCGCGCAGC |
| 3781 | CGAACGACCG | AGCGCAGCGA | GTCAAGTGAAC | GAGGAAGCGG | AAGAGGCC | AATACGGAAA | CCGCCTCTCC |
| 3851 | CGCGCGCTTG | GCCGATTCAAT | TAATGCGAGCT | GGCACGACAG | GTFTCCCGAC | TGAAAGGGG | GCAGTGAGCC |

FIG._39D

74 / 154

3921 CAACGCAATT ATATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCGGGCTCGT
3991 ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TAGGCCAAGC
KpnI ~~~~~~
~~~~~  
4061 GCGCAATTAA CCCTCACTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCC CTCGAGGTCA TTCAATATGCT  
4131 TGAGAAGAGA GTCGGGATAG TCCAAAATA AACAAAGGTAA AGATTACCTG GTCAAAAGTG AAAACATCAG  
4201 TTAAAAGGTG GTATAAGTAA AATATCGGTAA ATAAAAGGTG GCCCAAAAGTG AAATTTACTC TTTCCTACTA  
4271 TTATAAAAT TGAGGATGTT TTGTCGGTAC TTTGATACTG CATTTTGTA TGAATTGTT TTAAAGTGT TTAAAGTGT  
4341 TTCGCGATTG GAAATGCAATCTGTT GAGTCGGTT TTAAGTGTGTT TGCTTTCGTT AATACAGAGG  
4411 GATTGTATA AGAAAATATCT TTAAAAAACCATATGCTAA TTTGACATAA TTTTGAGAA AAATATATAT  
KpnI ~~~~~~  
~~~~~  
ECORI ~~~~~~
~~~~~  
4481 TCAGGGGAAT TCCACAATGA ACAATAATA GATTAATAA GCTTGGCCCCC GTGGCAGCGA TGGGTATT  
4551 TTCTAGTAA ATTAAAGATA AACTTAGACT CAAACACATT ACAAAACAA CCCCTAAAGT CCTAAAGCCC  
4621 AAAGTGTAT GCACGATCCA TAGCAAGGCC AGCCCAACCC AACCCAAACCC AAGTCGAGCCA  
4691 ACTGGCAAAAT AGTCTCCACC CCCGGCACTA TCACCGTGTAG TTGTCGGCAC CACCGCACGT CTCGGCAGCCA  
4761 AAAAAGAAA AGAAAGAAA AAAAAGAAA AGAAAACAG CAGGTGGTCA CGGGTGTGG GGGCGGGAAA  
4831 AGCGAGGAGG ATCGGGAGCA GCGACGAGGC CGGGCCCTCC CTCGGTTC AAGAAAGC CCCCCATCGC  
4901 CACTATAC ATACCCCCC CTCTCCTCCC ATCCCCAA CCCTACCA ACCACCCA CCACCTCCCTC  
4971 CCCCTCGCT GCGGGACGAC GAGCTCCTCC CCCCTCCCC TCCGCCGCC CCGTAACCA CCCCGCCCT  
5041 CTCCTCTTC TTCTCCTCGTT TTCTGGTCTC TCTGGTCTC GATCTTGGC CTGGTAGTT TGGTGGGG  
5111 AGAGCGGCTT CGTCGCCAG ATCGGTGCGC GGGAGGGGGC GCTGGCGTCT CGGGCGTGA  
BamHI ~~~~~~  
~~~~~  
5181 GTCGGGCCGG ATCCCTCGGG GGAATGGGGC TCTCGGATGT AGATCTCTCT TTTTCTCTCT TTGACAAAT GCAGCCTCGT
5251 GAATTGAAAT CCCTCAGCAT TGTTCATCGG TAGTTTCTCT TTTCATGATT TGTGACAAAT GCAGCCTCGT
5321 GCGGAGCTT TTGGTAGC

FIG._39E

75 / 154

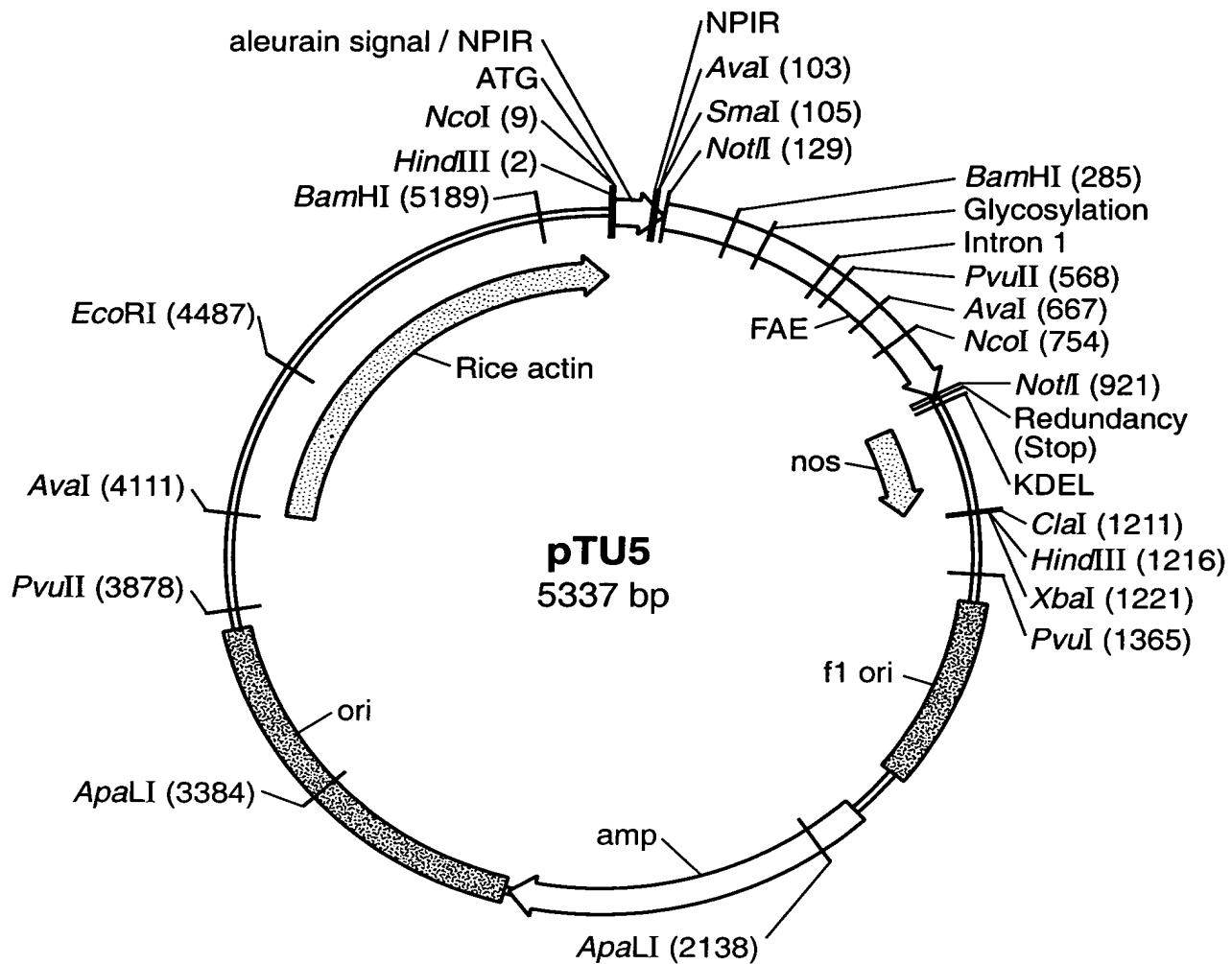


FIG._40A

76 / 154

HindIII NcoI

1 AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT
TTCGAATGGT ACCGGGTGCG GGCGCAGGAG GAGGACCGCG AGCGGCACGA
51 GGCCACGGCC GCCGTCGCCG TCGCCTCCTC CTCCTCCTTC GCCGACTCCA
CCGGTGCCGG CGGCAGCGGC AGCGGAGGAG GAGGAGGAAG CGGCTGAGGT

SmaI

101 ACCCGGGCCG GCCCGTCACC GACCGCGCGG CCGCCTCCAC GCAGGGCATC
TGGGCCCAGGC CGGGCAGTGG CTGGCGCGCC GGCGGAGGTG CGTCCCCTAG
151 TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC
AGGCTTCTGG AGATGTCGGC AAATCAGCTT TACCGGTGAT AGAGGGTTCG
201 TGCCTACGCC GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA
ACGGATGCGG CTGGACACGT TGTAAGGCAG CTGATAATAG TTCCCTCTCT

BamHI

251 AAATTTACAA TTCTCAAACG GACATTAACG GATGGATCCT CCGCGACGAC
TTTAAATGTT AAGAGTTTGA CTGTAATTGC CTACCTAGGA GGCGCTGCTG
301 AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACTGGTA GTGATACGAA
TCGTCGTTTC TTTATTAGTG GCAGAAGGCA CCGTGACCAT CACTATGCTT
351 TCTACAACTC GATACTAACT ACACCCTCAC GCCTTCGAC ACCCTACCAC
AGATGTTGAG CTATGATTGA TGTGGGAGTG CGGAAAGCTG TGGGATGGTG
401 AATGCAACGG TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC
TTACGTTGCC AACACTTCAT GTGCCACCTA TAATATAACC TACCCAGAGG
451 GTCCAGGACC AAGTCGAGTC GCTTGTAAA CAGCAGGTAA GCCAGTATCC
CAGGTCTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT CGGTCACTAGG
501 GGACTACGCG CTGACCGTG A CGGCCACKC CCTCGGGCGC TCCCTGGCGG
CCTGATGCGC GACTGGCACT GGCGGGTGMG GGAGCCGCGG AGGGACCGCC

PvuII

551 CACTCACTGC CGCCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC
GTGAGTGACG CGGGGTCGAC AGACGCTGTA TGCTGTTGTA GGCGGACATG
601 ACCTTCGGCG AACCGCGCAG CGGCAATCAG GCCTTCGCGT CGTACATGAA
TGGAAGCCGC TTGGCGCGTC GCCGTTAGTC CGGAAGCGCA GCATGTACTT

AvaI

651 CGATGCCTTC CAAGCCTCGA GCCCAGATA GACGCAGTAT TTCCGGGTCA
GCTACGGAAG GTTCGGAGCT CGGGTCTATG CTGCGTCATA AAGGCCAGT

FIG._40B

77 / 154

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC
GAGTACGGTT GCTGCCGTAG GGTTTGGACG GGGGCCACCT CGTCCCCATG

NcoI

~~~~~

751 GCCCATGGCG GTGTAGAGTA CTGGAGCGTT GATCCTTACA GCGCCCAGAA  
CGGGTACCGC CACATCTCAT GACCTCGCAA CTAGGAATGT CGCGGGTCTT

801 CACATTGTC TGCACGGGG ATGAAGTGCA GTGCTGTGAG GCCCAGGGCG  
GTGTAAACAG ACGTGACCCC TACTTCACGT CACGACACTC CGGGTCCCGC

851 GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGC  
CTGTCCCACA CTTATTACGC GTGTGCTGAA TAAAACCCTA CTGCTCGCCG

**NotI**

~~~~~

901 GCATGCACCT GGCCGGTCGC GGCCGCGGAA CCACTGAAGG ATGAGCTGTA
CGTACGTGGA CGGGCCAGCG CGGGCGCCTT GGTGACTTCC TACTCGACAT

951 AAGAACGAGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AAGATTGAAT
TTCTCGTCT AGCAAGTTG TAAACCGTTA TTTCAAAGAA TTCTAACTTA

1001 CCTGTTGCCG GTCTTGCAT GATTATCATA TAATTTCTGT TGAATTACGT
GGACAAACGGC CAGAACGCTA CTAATAGTAT ATAAAGACA ACTTAATGCA

1051 TAAGCATGTA ATAATTAACA TGTAATGCAT GACGTTATTT ATGAGATGGG
ATTCGTACAT TATTAATTGT ACATTACGTA CTGCAATAAA TACTCTACCC

1101 TTTTATGAT TAGAGTCCCG CAATTATACA TTTAATACGC GATAGAAAAC
AAAAATACTA ATCTCAGGGC GTTAATATGT AAATTATGCG CTATCTTTG

1151 AAAATATAGC GCGAAACTA GGATAAAATTA TCGCGCGCGG TGTCACTAT
TTTATATCG CGCGTTGAT CCTATTTAAT AGCGCGCGCC ACAGTAGATA

XbaI

~~~~~

**Clal HindIII**

~~~~~

1201 GTTACTAGAT CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTGCC
CAATGATCTA GCTATTGAA GATCTCGCCG GCCACCTCGA GGTTAAGCGG

1251 CTATAGTGAG TCGTATTACG CGCGCTCACT GGCGTCGTT TTACAACGTC
GATATCACTC AGCATAATGC GCGCGAGTGA CGGGCAGCAA AATGTTGAG

1301 GTGACTGGGA AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT
CACTGACCCCT TTTGGGACCG CAATGGGTTG AATTAGCGGA ACGTCGTGTA

PvuII

~~~~~

1351 CCCCCTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC  
GGGGGAAAGC GGTCGACCGC ATTATCGCTT CTCCGGGCGT GGCTAGCGGG

1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCCTGTAGCG  
AAGGGTTGTC AACGCGTCGG ACTTACCGCT TACCTCGCGC GGGACATCGC

**FIG.\_40C**

1451 GCGCATTAAAG CGCGGCGGGT GTGGTGGTTA CGCGCAGCGT GACCGCTACA  
CGCGTAATTC GCGCCGCCA CACCACCAAT GCGCGTCGCA CTGGCGATGT

1501 CTTGCCAGCG CCCTAGCGCC CGCTCCTTTC GCTTTCTTCC CTTCCTTCT  
GAACGGTCGC GGGATCGCGG GCGAGGAAAG CGAAAGAAGG GAAGGAAAGA

1551 CGCCACGTTG GCCGGCTTTC CCCGTCAAGC TCTAAATCGG GGGCTCCCTT  
GCGGTGCAAG CGGCCGAAAG GGGCAGTTCG AGATTAGCC CCCGAGGGAA

1601 TAGGGTTCCG ATTTAGTGCT TTACGGCACC TCGACCCCAA AAAACTTGAT  
ATCCCAAGGC TAAATCACGA AATGCCGTGG AGCTGGGTT TTTGAACTA

1651 TAGGGTGATG GTTCACGTAG TGGGCCATCG CCCTGATAGA CGGTTTTCG  
ATCCCACTAC CAAGTGCATC ACCCGGTAGC GGGACTATCT GCCAAAAAGC

1701 CCCTTGACG TTGGAGTCCA CGTTCTTAA TAGTGGACTC TTGTTCCAAA  
GGGAAACTGC AACCTCAGGT GCAAGAAATT ATCACCTGAG AACAAAGGTT

1751 CTGGAACAAAC ACTCAACCCCT ATCTCGGTCT ATTCTTTGA TTTATAAGGG  
GACCTTGTG TGAGTTGGGA TAGAGCCAGA TAAGAAAACT AAATATTCCC

1801 ATTTTGCCGA TTTCGGCCTA TTGGTTAAAA AATGAGCTGA TTTAACAAAA  
TAAAACGGCT AAAGCCGGAT AACCAATT TTACTCGACT AAATTGTTT

1851 ATTTAACGCG AATTTAACAA AAATATTAAC GCTTACAATT TAGGTGGCAC  
TAAATTGCGC TTAAAATTGT TTTATAATTG CGAATGTTAA ATCCACCGTG

1901 TTTTCGGGGA AATGTGCGCG GAACCCCTAT TTGTTTATTT TTCTAAATAC  
AAAAGCCCCT TTACACGCGC CTTGGGGATA AACAAATAAA AAGATTATG

1951 ATTCAAATAT GTATCCGCTC ATGAGACAAT AACCTGATA AATGCTCAA  
TAAGTTATA CATAGGCGAG TACTCTGTTA TTGGGACTAT TTACGAAGTT

2001 TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTCC GTGTCGCCCT  
ATTATAACTT TTTCCCTCTC ATACTCATAA GTTGTAAAGG CACAGCGGGA

2051 TATTCCTTT TTTGCGGCAT TTTGCCTTCC TGTTTTGCT CACCCAGAAA  
ATAAGGGAAA AACGCCGTA AACCGGAAGG ACAAAACGA GTGGGTCTTT

**ApaLI**

2101 CGCTGGTGAAGTAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT  
GCGACCACTT TCATTTCTA CGACTCTAG TCAACCCACG TGCTACCCA

2151 TACATCGAAC TGGATCTCAA CAGCGGTAAAG ATCCTTGAGA GTTTCGCCC  
ATGTAGCTTG ACCTAGAGTT GTCGCCATTG TAGGAACTCT CAAAAGCGGG

2201 CGAAGAACGT TTTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG  
GCTTCTTGCA AAAGGTTACT ACTCGTGAAA ATTTCAAGAC GATACACCGC

2251 CGGTATTATC CCGTATTGAC GCCGGGCAAG AGCAACTCGG TCGCCGCATA  
GCCATAATAG GGCATAACTG CGGCCCGTTC TCGTTGAGCC AGCGGCGTAT

2301 CACTATTCTC AGAATGACTT GGTTGAGTAC TCACCAAGTCA CAGAAAAGCA  
GTGATAAGAG TCTTACTGAA CCAACTCATG AGTGGTCAGT GTCTTTCGT

79 / 154

2351 TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA  
AGAATGCCTA CCGTACTGTC ATTCTCTTAA TACGTCACGA CGGTATTGGT

2401 TGAGTGATAA CACTCGGGCC AACTTACTTC TGACAAACGAT CGGAGGACCG  
ACTCACTATT GTGACGCCGG TTGAATGAAG ACTGTTGCTA GCCTCCTGGC

2451 AAGGAGCTAA CCGCTTTTT GCACAAACATG GGGGATCATG TAACTCGCCT  
TTCCTCGATT GGCGAAAAAA CGTGTGTTAC CCCCTAGTAC ATTGAGCGGA

2501 TGATCGTTGG GAACCGGAGC TGAATGAAGC CATAACCAAC GACGAGCGTG  
ACTAGCAACC CTTGGCCTCG ACTTACTTCG GTATGGTTG CTGCTCGCAC

2551 ACACCACGAT GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAAC  
TGTGGTGCTA CGGACATCGT TACCGTTGTT GCAACCGCGTT TGATAATTGA

2601 GGCAGAACTAC TTACTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA  
CCGCTTGATG AATGAGATCG AAGGGCCGTT GTTAATTATC TGACCTACCT

2651 GGCAGATAAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CGGGCTGGCT  
CCGCCTATTT CAACGTCCTG GTGAAGACGC GAGCCGGGAA GGCGACCGA

2701 GGTTTATTGC TGATAAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC  
CCAAATAACG ACTATTTAGA CCTCGGCCAC TCGCACCCAG AGCGCCATAG

2751 ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA  
TAACGTCGTG ACCCCGGTCT ACCATTGGG AGGGCATAGC ATCAATAGAT

2801 CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG  
GTGCTGCCCG TCAGTCCGTT GATACCTACT TGCTTATCT GTCTAGCGAC

2851 AGATAGGTGC CTCACTGATT AAGCATTGGT AACTGTCAGA CCAAGTTTAC  
TCTATCCACG GAGTGACTIONA TTCTGTAACCA TTGACAGTCT GGTCAAATG

2901 TCATATATAC TTTAGATTGA TTTAAAACCTT CATTTTTAAT TTAAAAGGAT  
AGTATATATG AAATCTAACT AAATTTGAA GTAAAAAATTA AATTTTCCTA

2951 CTAGGTGAAG ATCCTTTTG ATAATCTCAT GACCAAAATC CCTTAACGTG  
GATCCACTTC TAGGAAAAAC TATTAGAGTA CTGGTTTTAG GGAATTGCAC

3001 AGTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT  
TCAAAAGCAA GGTGACTCGC AGTCTGGGGC ATCTTTCTA GTTCCTAGA

3051 TCTTGAGATC CTTTTTTCT GCGCGTAATC TGCTGTTGC AAACAAAAAA  
AGAACTCTAG GAAAAAAAGA CGCGCATTAG ACGACGAACG TTTGTTTTT

3101 ACCACCGCTA CCAGCGGTGG TTTGTTGCC GGATCAAGAG CTACCAACTC  
TGGTGGCGAT GGTCGCCACC AAACAAACGG CCTAGTTCTC GATGGTTGAG

3151 TTTTCCGAA GGTAACTGGC TTCAGCAGAG CGCAGATACC AAATACTGTC  
AAAAAGGCTT CCATTGACCG AAGTCGTCTC GCGTCTATGG TTTATGACAG

3201 CTTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAACT CTGTAGCACC  
GAAGATCACA TCGGCATCAA TCCGGTGGTG AAGTTCTTGA GACATCGTGG

3251 GCCTACATAC CTCGCTCTGC TAATCCTGTT ACCAGTGGCT GCTGCCAGTG  
CGGATGTATG GAGCGAGACG ATTAGGACAA TGGTCACCGA CGACGGTCAC

FIG.\_40E

80 / 154

3301 GCGATAAGTC GTGTCTTACC GGGTTGGACT CAAGACGATA GTTACCGGAT  
CGCTATTCAG CACAGAATGG CCCAACCTGA GTTCTGCTAT CAATGGCCTA

**Apa**LI

~~~~~

3351 AAGGCGCAGC GGTCGGGCTG AACGGGGGTT TCGTGCACAC AGCCCAGCTT
TTCCCGCTCG CCAGCCCGAC TTGCCCCCA AGCACGTGTG TCGGGTCGAA

3401 GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG
CCTCGCTTGC TGGATGTGGC TTGACTCTAT GGATGTCGCA CTCGATACTC

3451 AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CGGACAGGTA TCCGGTAAGC
TTTCGCGGTG CGAAGGGCTT CCCTCTTCC GCCTGTCCAT AGGCCATTG

3501 GGCAGGGTCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG GGGGAAACGC
CCGTCCCAGC CTTGTCCCT CGCGTGCTCC CTCGAAGGTC CCCCTTGCG

3551 CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC
GACCATAGAA ATATCAGGAC AGCCCAAAGC GGTGGAGACT GAACTCGCAG

3601 GATTTTGTG ATGCTCGTCA GGGGGGCGGA GCCTATGGAA AAACGCCAGC
CTAAAAAACAC TACGAGCAGT CCCCCCGCCT CGGATAACCTT TTTGCGGTG

3651 AACGCGGCCT TTTTACGGTT CCTGGCCTTT TGCTGGCCTT TTGCTCACAT
TTGCGCCGGA AAAATGCCAA GGACCGGAAA ACGACCGGAA AACGAGTGT

3701 GTTCTTCCT GCGTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCCT
CAAGAAAGGA CGCAATAGGG GACTAAGACA CCTATTGGCA TAATGGCGGA

3751 TTGAGTGAGC TGATACCGCT CGCCGCAGCC GAACGACCGA GCGCAGCGAG
AACTCACTCG ACTATGGCGA GCGCGTGTGG CTTGCTGGCT CGCGTCGCTC

3801 TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAAC CGCCTCTCCC
AGTCACTCGC TCCTTCGCCT TCTCGCGGGT TATGCGTTG GCGGAGAGGG

PvuII

~~~~~

3851 CGCGCGTTGG CCGATTCTATT AATGCAGCTG GCACGACAGG TTTCCGACT  
GCGCGCAACC GGCTAAGTAA TTACGTCGAC CGTGTGTCC AAAGGGCTGA

3901 GGAAAGCGGG CAGTGAGCGC AACGCAATTAA ATGTGAGTTA GCTCACTCAT  
CCTTTCGCCCG GTCACTCGCG TTGCGTTAAT TACACTCAAT CGAGTGAGTA

3951 TAGGCACCCCC AGGCTTTACA CTTTATGCTT CCGGCTCGTA TGTTGTGTGG  
ATCCGTGGGG TCCGAAATGT GAAATACGAA GGCGAGCAT ACAACACACC

4001 AATTGTGAGC GGATAACAAAT TTCACACAGG AAACAGCTAT GACCATGATT  
TTAACACTCG CCTATTGTTA AAGTGTGTCC TTTGTCGATA CTGGTACTAA

4051 ACGCCAAGCG CGCAATTAAAC CCTCACTAAA GGGAACAAAAA GCTGGGTACC  
TGCGGTTCGC GCGTTAATTG GGAGTGATTT CCCTTGTGTT CGACCCATGG

**Ava**I

~~~~~

4101 GGGCCCCCCCCC TCGAGGGTCAT TCATATGCTT GAGAAGAGAG TCGGGATAGT
CCCGGGGGGG AGCTCCAGTA AGTATACGAA CTCTTCTCTC AGCCCTATCA

FIG._40F

81 / 154

4151 CCAAAATAAA ACAAAAGGTAA GATTACCTGG TCAAAAGTGA AAACATCAGT
GGTTTATTT TGTTTCCATT CTAATGGACC AGTTTCACT TTTGTAGTCA

4201 TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAAGGTGG CCCAAAGTGA
ATTTTCCACC ATATTCAATT TATAGCCATT ATTTTCCACC GGGTTTCACT

4251 AATTACTCT TTTCTACTAT TATAAAAATT GAGGATGTTT TGTCGGTACT
TTAAATGAGA AAAGATGATA ATATTTTAA CTCCTACAAA ACAGCCATGA

4301 TTGATACGTC ATTTTGTAT GAATTGGTTT TTAAGTTTAT TCGCGATTTG
AACTATGCAG TAAAAACATA CTTAACAAA AATTCAAATA AGCGCTAAC

4351 GAAATGCATA TCTGTATTTG AGTCGGTTT TAAGTCGTT GCTTTGTAA
CTTTACGTAT AGACATAAAC TCAGCCAAA ATTCAAGCAA CGAAAACATT

4401 ATACAGAGGG ATTTGTATAA GAAATATCTT TAAAAAACCC ATATGCTAAT
TATGCTCCC TAAACATATT CTTTATAGAA ATTTTTGGG TATACGATTA

ECORI

~~~~~

4451 TTGACATAAT TTTGAGAAA AATATATATT CAGGCGAATT CCACAATGAA  
AACTGTATTA AAAACTCTTT TTATATATAA GTCCGCTTAA GGTGTTACTT  
  
4501 CAATAATAAG ATTAAAATAG CTTGCCCG TTGCAGCGAT GGGTATTTT  
GTTATTATTC TAATTTTATC GAACGGGGC AACGTCGCTA CCCATAAAA  
  
4551 TCTAGTAAAA TAAAAGATAA ACTTAGACTC AAAACATTAA CAAAAACAAAC  
AGATCATTAA ATTTCCTATT TGAATCTGAG TTTGTAAAT GTTTTGTGTT  
  
4601 CCCTAAAGTC CTAAAGCCCA AAGTGTATG CACGATCCAT AGCAAGCCCA  
GGGATTCAG GATTCGGGT TTCACGATAC GTGCTAGGTA TCGTTCGGGT  
  
4651 GCCCAACCCA ACCCAACCCA ACCCACCCCA GTGCAGCCAA CTGGCAAATA  
CGGGTTGGGT TGGGTTGGGT TGGGTGGGT CACGTCGGTT GACC GTTTAT  
  
4701 GTCTCCACCC CGGGCACTAT CACCGTGAGT TGTCCGCACC ACCGCACGTC  
CAGAGGTGGG GGCGTGATA GTGGCACTCA ACAGGCGTGG TGGCGTGCAG  
  
4751 TCGCAGCCAA AAAAAAAA AGAAAGAAA AAAAGAAAA GAAAACAGC  
AGCGTCGGTT TTTTTTTTT TCTTCTTTT TTTTCTTTT CTTTTGTGCG  
  
4801 AGGTGGGTCC GGGTCGTGGG GGCCGGAAAA GCGAGGAGGA TCGCGAGCAG  
TCCACCCAGG CCCAGCACCC CGGGCCTTT CGCTCCTCCT AGCGCTCGTC  
  
4851 CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC  
GCTGCTCCGG GCCGGGAGGG AGGCGAAGGT TTCTTGCAG GGGTAGCGG  
  
4901 ACTATATACA TACCCCCCCC TCTCCTCCCA TCCCCCAAC CCTACCAACCA  
TGATATATGT ATGGGGGGGG AGAGGAGGGT AGGGGGGTTG GGATGGTGGT  
  
4951 CCACCACAC CACCTCCTCC CCCCTCGCTG CGGGACGACG AGCTCCTCCC  
GGTGGTGGTG GTGGAGGAGG GGGGAGCGAC GGCCTGCTGC TCGAGGAGGG  
  
5001 CCCTCCCCCT CCGCCGCCGC CGGTAACCAC CCCGCCCTC TCCTCTTCT  
GGGAGGGGGA GGCAGCGGGCG GCCATTGGTG GGGCGGGGAG AGGAGAAAGA

FIG.\_40G

82 / 154

5051 TTCTCCGTTT TTTTTTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT  
AAGAGGCAAA AAAAAAAGCA GAGCCAGAGC TAGAAACCAG AACCATCAAA

5101 GGGTGGGCGA GAGCGGCTTC GTGCCCCAGA TCGGTGCGCG GGAGGGGCGG  
CCCACCCGCT CTCGCCGAAG CAGCGGGTCT AGCCACGCGC CCTCCCCGCC

BamHI

~~~~~

5151 GATCTCGCGG CTGGCGTCTC CGGGCGTGAG TCGGCCCCGA TCCTCGCGGG
CTAGAGCGCC GACCGCAGAG GCCCGCACTC AGCCGGGCCT AGGAGCGCCC

5201 GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTTGTGGTAG
CTTACCCCGA GAGCCTACAT CTAGAAGAAA GAAAGAAGAA AAACACCATC

5251 AATTGAAATC CCTCAGCATT GTTCATCGGT AGTTTTCTT TTCATGATTT
TTAAACTTAG GGAGTCGTAA CAAGTAGCCA TCAAAAAGAA AAGTACTAAA

5301 GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGC
CACTGTTAC GTCGGAGCAC GCCTCGAAAA AACATCG

FIG._40H

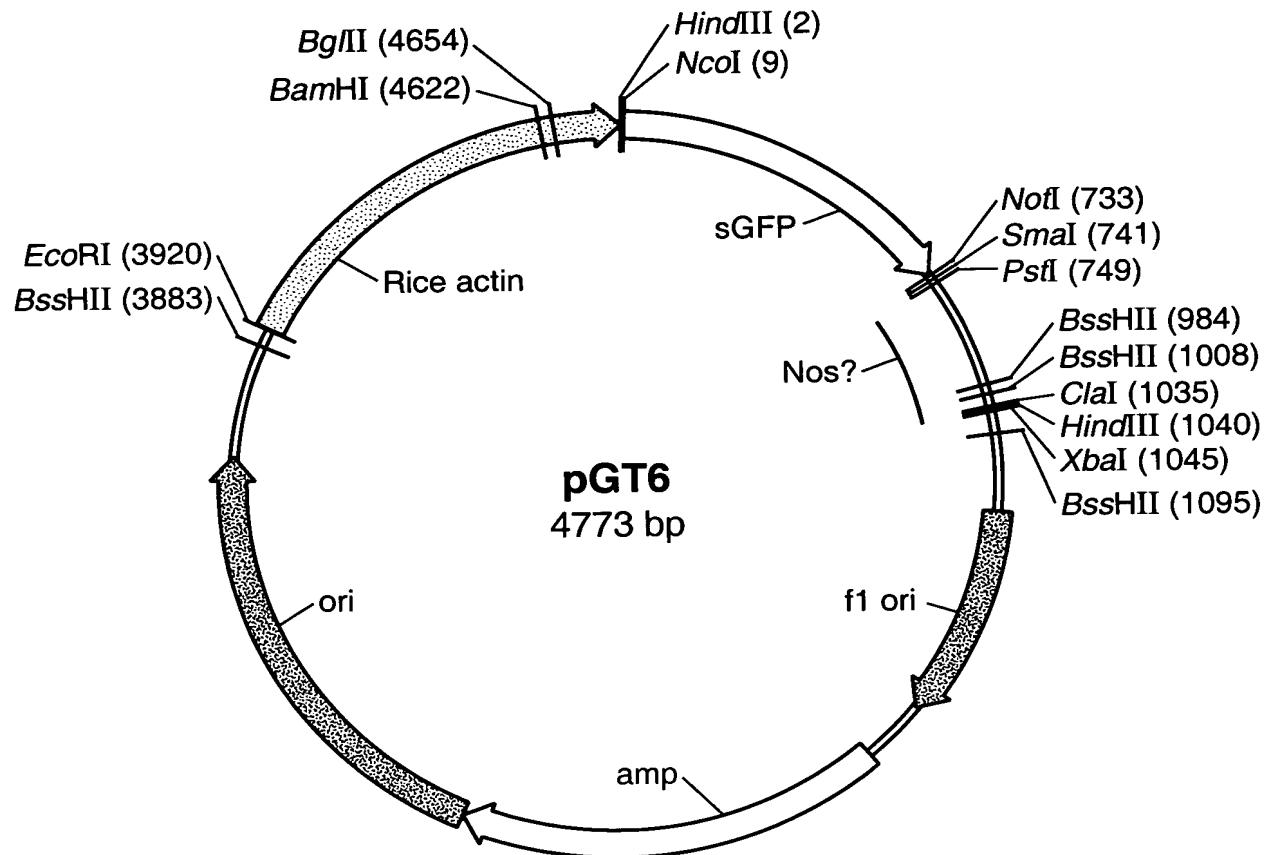


FIG._41A

| HindIII NcoI | |
|--------------|--|
| 1 | AAGCTTACCA TGGTGGCAA GGGCGAGGAG CTGTTCACCG GGGTGGGCC CATCCGGC GAGCTGGTC GAGCTGGAG TCGGATGGT ACCACTCGTT CCCGTCCTC GACAAGTGGC CCCACCGG GTAGGACCA CTCGACCTGC |
| 71 | GCGACGTGAA CGGCCACAAG TTCAAGCGTGT CGGGCGAGGG CGAGGGCGAT GCCACCTACG GCAAGCTGAC CGCTGGCACTT GCGGGTGTTC AAGTCGACA GCGCGCTCC GCTCCCGCTA CGGTGGATGC CGTTCGACTG |
| 141 | CCTGAAGTTC ATCTGCACCA CGGGCAAGCT GCGCGTGGCC TGGCCCACCC TCGTACCC CTTCACCTAC GGACTTCAAG TAGACGTGGT GGCGGTCTGA CGGGCACGGG ACCGGGTGG AGCACTGGTG GAAGTGGATG |
| 211 | GGCGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAGC AGCACGACTT CTTCAAGTCC GCCATGCCCG CCGCACGTCGCA CGAACGTCGGC GATGGGGCTG GTGTACTTCG TCGTGTGAA GAAGTTCAGG CGGTACGGGC |
| 281 | AGGGCTACGT CCAGGGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC AAGACCCGG CCGAGGTGAA TTCCGATGCA GGTCCCTCGCG TGGTAGAAGA AGTTCCCTGT GCGGTGATG TTCTGGGGC GGCTCCACTT |
| 351 | GTTCGAGGGC GACACCCCTGG TGAACCCGCAT CGAGCTGAAG GGCATCGACT TCAAGGAGGA CGGCAACATC CAAGCTCCCG CTGTGGGACG ACTTGGGCTA GCTCGACTTC CCGTAGCTGA AGTTCCCTCT GCCGTTGTAG |
| 421 | CTGGGCACCA AGCTGGAGTA CAACTAACAC AGCCACAAACG TCTATATCAT GGCCGACAAAG CAGAAGAACG GACCCCGTGT TCGACCTCAT GTTGTATGTTG TCGGTGTGC AGATATAGTA CCGGCTGTTT GCCTTCTTGC |
| 491 | GCATCAAGGT GAACTTCAG ATCCGCCACA ACATCGAGGA CGGGAGCGTG CAGCTGCCG ACCACTACCA CGTAGTTCCA CTTGAAGTTCTAGGCGGTGT TGTAGCTCTT GCCGTCGCC ACCTGAGGCC TGGTGATGGT |
| 561 | GCAGAACACC CCCATCGGGC ACGGCCCGT GCTGCTGCC GACAACACT ACCTGAGCAC CCAGTCGCC CGTCCTGTGG GGGTAGCCG TGGCGGGCA CGACGGGGG CTGTTGGTA TGGACTCGTG GGTCAAGGGG |
| 631 | CTGAGCAAG ACCCCAACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT CGTGACCGGC GCGGGATCA GACTCGTTTC TGGGGTTGCT CTTCGGGCTA GTGTACCAAGG AGCACCTCAA GCACACTGGGG CGGCCCTAGT |

| | | SmaI | NotI | PstI | XbaI | BssHII | BssHII | Clal HindIII | |
|------|---|---|---|---|---|---|---|---|---|
| 701 | CTCACGGCAT GGACGGCTG TACAAGTAAA GAGTGGCCGTA CCTGCTCGAC | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| | GGGGCCGGCC GGGCTGCAGG GAAACCACTG AAGGATGAGC | GGGGCCGGCC GGGCTGCAGG GAAACCACTG AAGGATGAGC | GGGGCCGGCC GGGCTGCAGG GAAACCACTG AAGGATGAGC | GGGGCCGGCC GGGCTGCAGG GAAACCACTG AAGGATGAGC | GGGGCCGGCC GGGCTGCAGG GAAACCACTG AAGGATGAGC | GGGGCCGGCC GGGCTGCAGG GAAACCACTG AAGGATGAGC | GGGGCCGGCC GGGCTGCAGG GAAACCACTG AAGGATGAGC | GGGGCCGGCC GGGCTGCAGG GAAACCACTG AAGGATGAGC | GGGGCCGGCC GGGCTGCAGG GAAACCACTG AAGGATGAGC |
| 771 | TGTAAGAAC CAGATCGTT CAATAAAGTT TCTTAAGATT GAATCCGTG | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| | ACATTTCTTC GTCTAGCAAG TTGTAAACC GTTATTCAA AGAATTCAA | ACATTTCTTC GTCTAGCAAG TTGTAAACC GTTATTCAA AGAATTCAA | ACATTTCTTC GTCTAGCAAG TTGTAAACC GTTATTCAA AGAATTCAA | ACATTTCTTC GTCTAGCAAG TTGTAAACC GTTATTCAA AGAATTCAA | ACATTTCTTC GTCTAGCAAG TTGTAAACC GTTATTCAA AGAATTCAA | ACATTTCTTC GTCTAGCAAG TTGTAAACC GTTATTCAA AGAATTCAA | ACATTTCTTC GTCTAGCAAG TTGTAAACC GTTATTCAA AGAATTCAA | ACATTTCTTC GTCTAGCAAG TTGTAAACC GTTATTCAA AGAATTCAA | ACATTTCTTC GTCTAGCAAG TTGTAAACC GTTATTCAA AGAATTCAA |
| 841 | CGATGATTAT CATAATAATT CTGTTGAATT ACGTTAAGCA TGTTATAATT AACATGTAAT GCATGACGTT | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| | GCTACTAATA GTATTTAAA GACAACCTAA TGCAATTCTGT ACATTATAA TTGTACATTA CGTACTGCAA | GCTACTAATA GTATTTAAA GACAACCTAA TGCAATTCTGT ACATTATAA TTGTACATTA CGTACTGCAA | GCTACTAATA GTATTTAAA GACAACCTAA TGCAATTCTGT ACATTATAA TTGTACATTA CGTACTGCAA | GCTACTAATA GTATTTAAA GACAACCTAA TGCAATTCTGT ACATTATAA TTGTACATTA CGTACTGCAA | GCTACTAATA GTATTTAAA GACAACCTAA TGCAATTCTGT ACATTATAA TTGTACATTA CGTACTGCAA | GCTACTAATA GTATTTAAA GACAACCTAA TGCAATTCTGT ACATTATAA TTGTACATTA CGTACTGCAA | GCTACTAATA GTATTTAAA GACAACCTAA TGCAATTCTGT ACATTATAA TTGTACATTA CGTACTGCAA | GCTACTAATA GTATTTAAA GACAACCTAA TGCAATTCTGT ACATTATAA TTGTACATTA CGTACTGCAA | GCTACTAATA GTATTTAAA GACAACCTAA TGCAATTCTGT ACATTATAA TTGTACATTA CGTACTGCAA |
| 911 | ATTATGAGA TGGTTTTTTA TGATTAGAGT CCCGCAATTAA TACATTAAAT AGCGGATAGA AAACAAAATAA | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| | TAATAACTCT ACCCAAAAT ACTAATCTCA GGGCGTTAAAT ATGTAATTA TGGCCTATCT TTTGTTTTAT | TAATAACTCT ACCCAAAAT ACTAATCTCA GGGCGTTAAAT ATGTAATTA TGGCCTATCT TTTGTTTTAT | TAATAACTCT ACCCAAAAT ACTAATCTCA GGGCGTTAAAT ATGTAATTA TGGCCTATCT TTTGTTTTAT | TAATAACTCT ACCCAAAAT ACTAATCTCA GGGCGTTAAAT ATGTAATTA TGGCCTATCT TTTGTTTTAT | TAATAACTCT ACCCAAAAT ACTAATCTCA GGGCGTTAAAT ATGTAATTA TGGCCTATCT TTTGTTTTAT | TAATAACTCT ACCCAAAAT ACTAATCTCA GGGCGTTAAAT ATGTAATTA TGGCCTATCT TTTGTTTTAT | TAATAACTCT ACCCAAAAT ACTAATCTCA GGGCGTTAAAT ATGTAATTA TGGCCTATCT TTTGTTTTAT | TAATAACTCT ACCCAAAAT ACTAATCTCA GGGCGTTAAAT ATGTAATTA TGGCCTATCT TTTGTTTTAT | TAATAACTCT ACCCAAAAT ACTAATCTCA GGGCGTTAAAT ATGTAATTA TGGCCTATCT TTTGTTTTAT |
| 1051 | CGGCCGGTGG AGCTCCAATT CGCCCTATAG TGAGTCTGTT TACGGCGCGT | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| | GGCGGCCAAC TCGAGGTTAA GCGGGATATC ACTCAGGATA ATGCGCGGA | GGCGGCCAAC TCGAGGTTAA GCGGGATATC ACTCAGGATA ATGCGCGGA | GGCGGCCAAC TCGAGGTTAA GCGGGATATC ACTCAGGATA ATGCGCGGA | GGCGGCCAAC TCGAGGTTAA GCGGGATATC ACTCAGGATA ATGCGCGGA | GGCGGCCAAC TCGAGGTTAA GCGGGATATC ACTCAGGATA ATGCGCGGA | GGCGGCCAAC TCGAGGTTAA GCGGGATATC ACTCAGGATA ATGCGCGGA | GGCGGCCAAC TCGAGGTTAA GCGGGATATC ACTCAGGATA ATGCGCGGA | GGCGGCCAAC TCGAGGTTAA GCGGGATATC ACTCAGGATA ATGCGCGGA | GGCGGCCAAC TCGAGGTTAA GCGGGATATC ACTCAGGATA ATGCGCGGA |
| 1121 | CGTCGTGACT GGGAAAACCC TGGCGTTACC CAACTTAATC GCCTTGCGAGC | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| | GCAGGCACTGA CCCTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG | GCAGGCACTGA CCCTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG | GCAGGCACTGA CCCTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG | GCAGGCACTGA CCCTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG | GCAGGCACTGA CCCTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG | GCAGGCACTGA CCCTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG | GCAGGCACTGA CCCTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG | GCAGGCACTGA CCCTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG | GCAGGCACTGA CCCTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG |
| 1191 | GGCGTAATAG CGAAGAGGCC CGCACCGATC GCCCTTCCCA ACAGTTGCC | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| | CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAAGGGT TGTCACGG | CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAAGGGT TGTCACGG | CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAAGGGT TGTCACGG | CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAAGGGT TGTCACGG | CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAAGGGT TGTCACGG | CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAAGGGT TGTCACGG | CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAAGGGT TGTCACGG | CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAAGGGT TGTCACGG | CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAAGGGT TGTCACGG |
| 1261 | CGGCCCTGT AGGGGGCAT TAAGGGGGC GGGTGTGGTG GTTACCGGCC | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ | ~~~~~ |
| | GGGGGGGACA TCGCCGGGTG ATTTCGGCCG CCCACACAC CAATGGCGGT | GGGGGGGACA TCGCCGGGTG ATTTCGGCCG CCCACACAC CAATGGCGGT | GGGGGGGACA TCGCCGGGTG ATTTCGGCCG CCCACACAC CAATGGCGGT | GGGGGGGACA TCGCCGGGTG ATTTCGGCCG CCCACACAC CAATGGCGGT | GGGGGGGACA TCGCCGGGTG ATTTCGGCCG CCCACACAC CAATGGCGGT | GGGGGGGACA TCGCCGGGTG ATTTCGGCCG CCCACACAC CAATGGCGGT | GGGGGGGACA TCGCCGGGTG ATTTCGGCCG CCCACACAC CAATGGCGGT | GGGGGGGACA TCGCCGGGTG ATTTCGGCCG CCCACACAC CAATGGCGGT | GGGGGGGACA TCGCCGGGTG ATTTCGGCCG CCCACACAC CAATGGCGGT |

FIG._41C

1331 AGGCCCTAG CGCCCGCTCC TTTCGCTTTC TCCCTTCCT TTCTGCCAC GTTGCCTGGC TTTCCCGTC
TCGGGGATC GGGGGATC AAAGGAAAG AAGGAAGG AAGAGGGTGT CAAGGGCG AAAGGGCG
1401 AAGCTCTAA TCGGGGCTC CCTTTAGGGT TCCGATTAG TGCTTTACGG CACCTCGACC CCAAAACT
TTCGAGATT AGCCCCGAG GAAATCCCA AGGCTAATC ACGAAATGCCC GTGGAGCTGG GGTTTTGTGA
1471 TGATTAGGGT GATGGTCAC GTAGTGGGCC ATCGCCCTGA TAGACGGTT TTCGCCCTT GACGTTGGAG
ACTAATCCA CTACCAAGTG CATCACCCGG TAGCGGGACT ATCTGCCAA AAGGGAAA CTGCAACCTC
1541 TCCACGTTCT TAAATAGTGG ACTCTGGTC CAAACTGGAA CAACACTCAA CCCTATCTCG GTCTATTCTT
AGGTGCAAGA AATTATCACC TGAGAACAGG GTTGTGAGTT GGGATAGAGC CAGATAAGAA
1611 TTGATTATA AGGGATTGG CCGATTTCGG CCTATTGGTT AAAAATGAG CTGATTTAAC AAAAATTAA
AACTAAATAT TCCCTAAAC GGCTAAAGCC GGATAACCAA TTTTTACTC GACTAAATTG TTTTTAAATT
1681 CGCGAATTAAACAAATAT TAACGCTTAC AATTAGGTG GCACCTTTTCG GGGAAATGTTG CGCGGAACCC
GGCCTTAAAT TTGTTTTATA ATGGAAATG TAAATCCAC CGTGAAAGC CCCTTACAC GCGCCCTGGG
1751 CTATTTGTTT ATTTTCTAA ATACATCAA ATATGTTACCC GCTCATGAGA CAATAACCCCT GATAATGCT
GATAAACAA TAAAAGATT TATGTAAGTT TATACATAGG CGAGTACTCT GTTATTGGGA CTATTACGA
1821 TCAATAATAT TGAAAAGGA AGAGTATGAG TATTCAACAT TTCCGTTGTCG CCCTTATTCC CTTTTTGCG
AGTATTATA ACTTTTCCT TCTCATACTC ATAAGTTGTA AAGGCACAGC GGGATAAGG GAAAAAACGC
1891 GCATTTTGCC TTCCCTGTTT TGCTCACCCA GAAACGCTGG TGAAGTAAAGATAA AGATGCTGAA GATCAGTTGG
CGTAAACGG AAGGACAAA ACGAGTGGGT CTTGCGACCC ACTTTCATT TCTACGACTT CTAGTCAAAC
1961 GTGCACTGAGT GGGTACATC GAACTGGATC TCAACAGGG TAAGATCCTT GAGAGTTTC GCCCCGAAGA
CACGTGCTCA CCCAATGTAG CTTGACTAG AGTTGTGCC ATCTAGGAA CTCTCAAAAG CGGGCTTCT
2031 ACGTTTCACCA ATGATGAGCA CTTTAAAGT TCTGCTATGT GGCGGGTAT TATCCGTAT TGACGCCGGG
TGCAAAAGGT TACTACTCGT GAAAATTCA AGACGATACA CCGCGCCATA ATAGGGCATA ACTGGGCC
2101 CAAGAGCAAC TCGGTGCCG CATAACTAT TCTCAGAATG ACTTGGTTGA GTACTCACCA GTCACAGAAA
GTTCTCGTTG AGCCAGGGC GTATGTGATA AGAGTCTAC TGAACCAACT CATGAGTGGT CAGTGTCTTT

FIG._41D

2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA ACCATGAGTG ATAACACTGC
TCGTAGAATG CCTACCGGTAC TGTCAATTCTC TTAATAACGTC ACGACGGTAT TGGTACTCAC TATTGGACG

2241 GGCACACTTA CTTCTGACAA CGATCCGGAGG ACCGAAGGAG CTAACCGGTCT TTTCGACAA CATGGGGAT
CCGGTTGAAT GAAGACTGTT GCTAGCCTCC TGGCTTCCTC GATTGGGAA AAAACGTGTT GTACCCCTA

2311 CATGTAACTC GCCTGTGATCG TTGGAAACCG GAGCTGAATG AAGCCATACC AAACGACGAG CGTGACACCCA
GTACATTGAG CGGAACCTAGC AACCTTGGC CTCGACTTAC TTGGGTATGG TTGCTGCTC GCACTGTTGTT

2381 CGATGCCTGT AGCAATGGCA ACAACGTTGC GCAAACATTAACTGGCA AAACGACGAG CGTGACACCCA
GCTACGGACA TCGTGTACCGT TGTGCAACG CGTGTGATAA TTGACCCGCTT GATGAAATGAG ATCGAAGGGC

2451 GCAACAATTAA ATAGGACTGGA TGGAGGGGA TAAAGTTGCA GGACCACTTC TGCGCTCGGC CCTTCCGGCT
CGTGTGTTAAT TATCTGACCT ACCTCCGGCT ATTCAACGT CCTGGTGAAG ACGCGAGCCG GGAAGGGCA

2521 GGCTGGTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGGG TATCATGGCA GCACTGGGC
CCGACCAAAAT AACGACTATT TAGACCTCGG CCACTCGCAC CCAGAGGCC ATAGTAACGT CGTGACCCCG

2591 CAGATGGTA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGGAGTCAG GCAACTATGG ATGAAACGAAA
GTCTACCAT CGGGAGGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC CGTTGATAACCGTACTTGCTT

2661 TAGACAGATC GCTGAGATAG GTGCCTCACT GATTAAGCAT TGTTAACTGT CAGACCAAGT TTACTCAT
ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTGTA ACCATTGACA GTCTGGTTCA AATGAGTATA

2731 ATACTTTAGA TTGATTAAA ACTTCATTAA TTATTTAA GGATCTAGGT GAAGATCCTT TTTGATAATC
TATGAAATCT AACTAAATTG TGAAGTAAA ATTAATTT CCTAGATCCA CTTCTAGGAA AAACATTAG

2801 TCATGACCAA AATCCCTAA CGTGAGTTT CGTTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG
AGTACTGGTT TTAGGGAAATT GCACTCAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTT TCTAGTTCC

2871 ATCTTCTTGA GATCCTTTT TTCTGGCGT ATCTGGTGC TTGCAAACAA AAAAACACC GCTACCAAGGG
TAGAAGAACT CTAGGAAAAA AAGACGGGCA TTAGACGAG AACGTTTGTGTT TTTTGGTGG CGATGGTGC

2941 GTGGTTTGTG TGCCGGATCA AGAGCTACCA ACTCTTTTC CGAAGGTAAC TGGCTTCAGC AGAGGCCAGA
CACCAAACAA ACGGCCATAGT TCTCGATGGT TGAGAAAAG GCTTCCTAGT ACCGAAAGTCG TCTCGGGTCT

FIG._41E

3011 TACCAAATAC TGTCCCTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAG AACTCTGTAG CACCGCCTAC
ATGGTTATG ACAGGAAGAT CACATGGCA TCAATCCGGT GGTGAAGTTC TTGAGACATC GTGGGGATG

3081 ATACCTCGCT CTGCTTAATCC TGTACCACTG GGCTGGTGGC AGTGGCGATA AGTCGTGTCT TACCGGGTTG
TATGGCGA GACGATTAGG ACAATGGCA CCGACGACGG TCACCGCTAT TCAGCACAGA ATGGCCAAC

3151 GACTCAAGAC GATACTTAC GGATAAGGGC CAGGGGTCGG GCTGAACGGG GGTGTGTGC ACACAGGCCA
CTGAGTTCTG CTATCAATGG CCTATTCCGC GTCGCCAGCC CGACTTGCCC CCCAAGCACG TGTGTGGGT

3221 GCTTGGAGCG AACGACCTAC ACCGAACACTGA GATACTTACA GGCTGAGCTA TGAGAAAGCG CCACGCTTCC
CGAACCTCGC TTGCTGGATG TGGCTTGAAT CTATGGATGT CGCACTCGAT ACTCTTTCGC GGTGCGAAGG

3291 CGAAGGGAGA AAGGGGACA GGTATCCGGT AAGGGGAGG GTCGGAACAG GAGAGGCCAC GAGGGAGCTT
GCTTCCCTCT TTCCGCCCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGTG CTCTCGCGTG CTCCCTCGAA

3361 CCAGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTGGGT TTGCCCCACT CTGACTTGAG CGTCGATT
GGTCCCCCTT TGGGACCAT AGAAATATCA GACAGGCCA AAGGGTGGAA GACTGAACTC GCAGCTAAA

3431 TGTGATGCTC GTCAAGGGGG CGGAGCCTAT GGAAAACGCG CAGCAACGCC GCCTTTTAC GTTCTGGC
ACACTACGAG CAGTCCCCC GCCTCGGATA CCTTTTGTGCG GTCGTTGCC CGGAAAATG CCAAGGACCG

3501 CTTTTGCTGG CCTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT CTGTGGATAA CGGTATTAC
GAAAACGAC GAAAACGAG TGTACAAGAA AGGACGCAAT AGGGACTAA GACACCTATT GGCATAATGG

3571 GCCTTGTGAGT GAGCTGATAC CGCTGCCGC AGCCGAACGA CGAGGCCAG CGAGTCAGTG AGCGAGGAAG
CGGAAACTCA CTCGACTATG GCGAGGGCG TCGGCTTGTG GCTCAGTCAC TCGCTCCCTC

3641 CGGAAGGGCG CCCAATACGC AAACGGCCTC TCCCCGGG TTGGCCGATT CATTAAATGCA GCTGGCACGA
GCCTTCTCGC GGGTATGCG TTTGGGGAG AGGGGGCG AACCGGCTAA GTAATTACGT CGACCGTGT

3711 CAGGTTCCC GACTGGAAAG CGGGCAGTGA GCGCAACGCC ATAATGTCG GTTAGCTCAC TCATTAGGCA
GTCCAAGGG CTGACCTTTC GCGGTCACT CGCGTTGCGT TAATTACACT CAATCGAGTG AGTAATCCGT

3781 CCCCAGGGCT TACACTTTAT GCTTCCGGCT CGTATGTGT GTGGAATTGT GAGGGATAA CAATTACACA
GGGGTCCGAA ATGGAAATA CGAAGGGCGA GCATACACA CACCTTAACA CTCGGCTTATT GTTAAAGTGT

FIG.-41F

| | | ECORI | ECORI |
|------|--------------------------|---------------------------|---------------------------|
| 3851 | CAGGAAACAG GTCCTTTGTC | CTATGACCAT GATACTGGTA | AGCGGCCAAT TCCGGGGTAA |
| | | | TAACCCTCAC ATTGGAGTG |
| | | | TAAGGGAAC ATTCCCCTTG |
| | | | TTTTCGACCT |
| | | | |
| 3921 | ATTCCACAAT TAAGGTGTTA | GAACAAATAAT CTTCAATTAA | AAGATTAAAAA TTCTAAATT |
| | | | TAGCTTGGCCC ATCGAACGGG |
| | | | CCGTTGCAGC GGCAACGTCG |
| | | | GTGGGTATT CTACCCATAA |
| | | | TTTTCTAGTA AAAAGATCAT |
| 3991 | AAATAAAGA TTTATTTCCT | TAACCTTAGA ATTGAATCT | CTCAAAACAT GAGTTTGTA |
| | | | TTACAAAAC AATGTTTTG |
| | | | ACCCCTAAA TTGGGGATT |
| | | | GTCTAAAGC CAGGATTTCG |
| | | | CCAAAGTGCT GGTTACCGTT |
| 4061 | ATGCACGATC TACGTGCTAG | CATAGCAAGC GTATGTTCG | CCAGCCCAAC GGTGGGTTG |
| | | | CCAACCCAAC GGTTGGTTG |
| | | | CCAGTGCAGC GGTCACGTCG |
| | | | CAACTGGCAA GTTGACCGTT |
| 4131 | ATAGTCTCCA TATCAGAGGT | CCCCGGCAC GGGGCCGTG | TATCACCGTG ATAGTGGCAC |
| | | | AGTTGTCGC TCAACAGGG |
| | | | ACCAACCGAC TGGTGGGTG |
| | | | GTCTCGCAGC CAGAGGTTG |
| | | | CAAAAGAAAA GGGGGGGTG |
| 4201 | AAAAGAAAGA TTTCTTTCTT | AAAAGAAAC TTTCTTTCTT | AGCAGGTGGG TCGTCACCCC |
| | | | TGGTGGGTG AGGGCCAGCA |
| | | | CCCCGGGCT GGGGGGCT |
| | | | CCAAAGAAC GGTTCTT |
| | | | CCCCGGGCT GGGGGGCT |
| 4271 | GGATCGCGAG CCTAGCGCTC | CAGCGACGAG GTCTCGCTC | GCCCCCCT CGGGCCGGGA |
| | | | CCCTCCGCT GGGAGGGAA |
| | | | CCCTCCGCT GGGAGGGAA |
| | | | CCCTCCGCT GGGAGGGAA |
| 4341 | ACATAACCCCC TGTATGGGG | CCATCCCCC GGGAGGGAG | AACCCCTACCA TTGGGATGGT |
| | | | CCACCACCA GGTGGTGGTG |
| | | | CACCACCTCC GTGGTGGAGG |
| | | | CCCCCCC GGGGGGGG |
| 4411 | CTGCCGGAGC GACGGCCTGC | ACGAGGCTCCT TGCTCGAGGA | CGCCGGTAAC GGGGCCATTG |
| | | | CGCCGGCC GGGGGGGG |
| | | | CTCTCCTCTT GTGGGGGG |
| | | | CGAGAGGAGA |
| 4481 | TCTTCTCCG AGAAAGAGGC | TTTTTTTTTT AAAAAAAGA | CGTCTCGGTC GCAGAGCCAG |
| | | | TCTGATCTTG AGCTAGAAC |
| | | | GCCTGGTGT AAACCCACCC |
| | | | GCTCTGCCG |
| 4551 | TTCTGTCGCC AAGCAGCGGG | AGATGGGTGC TCTAGCCACG | GGGGAGGGG GCCCTCCCC |
| | | | CGGCTGGGTG GCCGACCGC |
| | | | CTCCGGGGT GAGTCGGCCC |
| | | | CTCAGCCGG CTCAGCCGG |

FIG.-41G

+

| | BamHI | BglII |
|------|---------------------------|----------------------------|
| 4621 | GGATCCCTCGC CCTAGGAGCG | GGGGAAATGGG CCCCTTACCC |
| 4691 | ATCCCTCAGC TAGGGAGTCG | ATTGGTTCAATC TAACAAGTAG |
| 4761 | TTTTGTAGG AAAAACATCC | TAG ATC |

FIG._41H

+

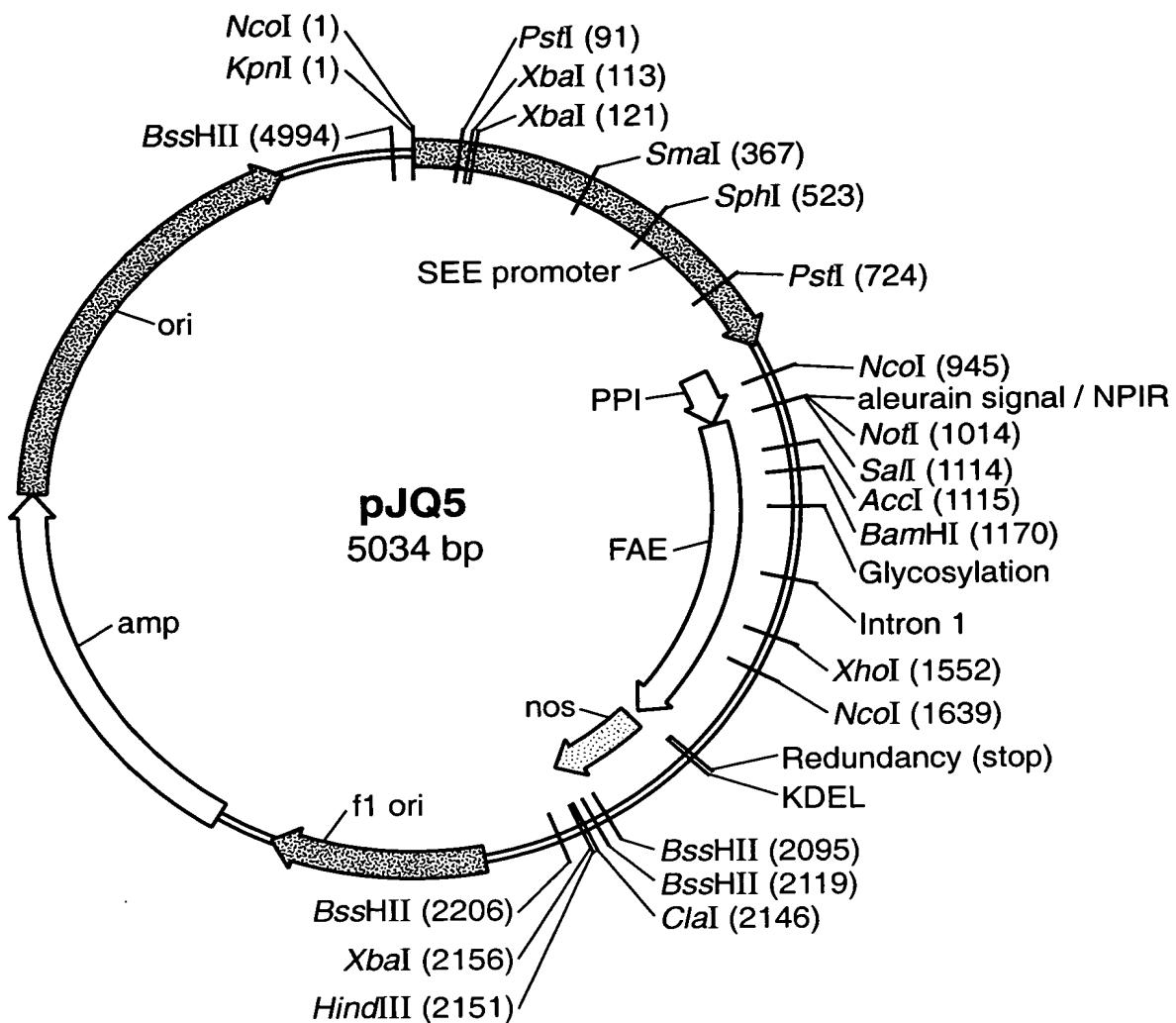


FIG._42A

NcoI
~~~~~  
KpnI

1 CATGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG  
GTACCCGGTC CATAATAA CCCTATAGAG TTCGTTTATT AGCTTATAG TGGTAACCGA TGTTATAGAC

PstI

71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC  
TCGAGGCTCA AGACTGACGT CAGACCTACT GGCACAAACA TAGATCTGA GATCTATCGT GTCGGGTGTGC  
141 ACCAACAGGA GTGCCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT  
TGGATGTCCCT CACGGTGTGA ACACCTGACA TCATCACAAAC CTCGGCTCG AGAAAGGATG GAGGACTGCA  
211 TGC CGCCCGT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCAACGGCTC CCAACAAAT ATCGTCCCCC  
ACGGGGCAA CAGGTAAGGT TGCCGTAGTG AGAGTTGGTT AGTGGCGAG GGTGTTTA TAGCAGGGGG  
281 ATGTCCTGGC GGAGAGAGAG TACATACATG CTGTCGGCC GTTTTTGTCT GAATCTCGCT TCCACTGGCC  
TACAGAACCG CCTCTCTCTC ATGTAATGTAC GACAGCGGG CAAAACAGA CTTAGAGCGA AGGTGACCGG

SmaI

351 AATCAGCTCA GCTCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA  
TTAGTCGAGT CGAGGGCCCT CGAGTCAGTA AGTCTAGGG TAGCAGCAGC AGTGGGGACC GCAGTACCT  
421 TGGAAAAGAA CCTCGTTGC TCGGATGAGT CAGCCATATC CCGGAACAGA GTACTGCAAG ATAACCCAAT  
ACCTTTCTT GGAGGCAACG AGCCTACTCA GTGGTATAG GGCTTGTCT CATGACGTTT CATTGGTTA

SphI

491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTGGG TTGGTGTGCTT CTTAATTGAC TTATTTTG  
AGTCTAAGGG GGTATCTCT TTCAATATCGT AGGAAAGCCC AAAACAAACC GAATTAAC TG AAATAAAAC  
561 TTGGAGTGTCA ATGGTGTGATT GTTGTGTAA ATGCCCAACC ATCTGAATAT CGAGACGGAT ATAGGGTGG  
AACCTCAACT TAGACTAAA CAACACATT TACGGGTGG TAGACTTATA GCTCTGCCTA TTATCGGACC

FIG.-42B



|      |                                                                                                                                                               |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1191 | CAAAGAAATA ATCACCGTCT TCCGGGCAC TGGTAGTGT ACGAATCTAC AACTCGATA CAACTACACC<br>GTTTCTTTAT TAGTGGCAGA AGGCACCGTG ACCATCACTA TGCTTAGATG TTGAGCTATG ATTGATGTGG     |
| 1261 | CTCACGCCCT TCGACACCCCT ACCACAATGTC AACGGTTGTG AAGTACACGG TGGATATTAT ATTGGATGGG<br>GAGTGGGAA AGCTGTGGGA TGGTGTACG TTGCCAACAC TTCAATGTGCC ACCTATAATA TAACCTACCC |
| 1331 | TCTCCGTCCA GGACCAAGTC GAGTCGCTTG TCAAACAGCA GGTTAGCCAG TATCCGGACT ACGGCCTGAC<br>AGAGGCAGGT CCTGGTTCACT CTCAGCGAAC AGTTTGTGCT CCAATGGTC ATAGGCCTGA TGGCGACTG   |
| 1401 | CGTGACCGGC CACKCCCTCG GGCCTCCCT GGGGGCACTC ACTGCGGCC AGCTGTCTGC GACATACGAC<br>GCACGTGGCG GTGMIIGGAGC CGGGGGAA CGGGCGTGGAG TGACGGGGG TCGACAGACG CTGTATGTG      |
| 1471 | AACATCCGCC TGTACACCTT CGCGAACCG CGCAGGGCA ATCAGGGCTT CGCGTCGTAC ATGAAACGATG<br>TTGTAGGGG ACATGTGGAA GCCGCTTGGC GCGTCGCCGT TAGTCCGGAA GCGCAGGATG TACTTGTAC     |
|      | XbaI                                                                                                                                                          |
| 1541 | CCTCCAAGC CTCGAGCCCA GATACGACGC AGTATTCCG GGTCACTCAT GCCAACGACG GCATCCAAA<br>GGAAGGTTCG GAGCTCGGGT CTATGCTGCG TCATAAAGG CCAAGTGAGTA CGGTGTCTGC CGTAGGGTT      |
|      | NcoI                                                                                                                                                          |
| 1611 | CCTGCCCG GTGGAGCAGG GGTACGCCA TGGGGGTGA GAGTACTGGA GCGTTGATCC TTACAGGCC<br>GGACGGGGC CACCTCGTCC CCATGGGGT ACCGCCACAT CTCAATGACCT CGCAACTAGG AATGTGGGG         |
| 1681 | CAGAACACAT TTGTTGTGCA TGGGATGAA GTGCAGTGT GTGAGGCCA GGGGGACAG GGTGTGAATA<br>GTCTTGTGTA AACAGACGTG ACCCCATTT CACGTCAAGA CACTCCGGT CCCGCTGTC CCACACTTAT         |
| 1751 | ATGGGCACAC GACTTATTG GGGATGACGA GCGGAGCTG TACATGGTGA TCAGTCATT CAGCCCTCCC<br>TACGGTGTG CTGAATAAAA CCCTACTGCT CGCCTCGGAC ATGTACCACT AGTCAGTAA GTCGGAGGG        |
| 1821 | GAGGTACCA GGAAAGATGG ATGTCCTGGA GAGGGGGCG CGTAACCACT GAAGGATGAG CTGTAAGAA<br>CTCACATGGT CCTTCTACC TACAGGACCT CTCCCCGGC GCATTTGGTA CTTCTACTC GACATTCTT         |

FIG.\_42D

95 / 154

|      |                                                                                                                                                              |              |
|------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| 1891 | GCAGATCGT CAAACATTG GCAATAAAGT TTCTTAAGAT TGAATCCTGT TGCCGGTCTT GCGATGATT<br>CGTCTAGCAA GTTGTAAAC CGTTATTCA AAGAATTCTA ACTTAGGACA ACGGCCAGAA CGCTACTAAT      | BssHII       |
| 1961 | TCATATAATT TCTGTTGAAT TAGTTAACG ATGTAATAAT TAAACATGTAAC TAGCATGACGT TATTATGAG<br>AGTATATTAA AGACAACTTA ATGCCAATTG TACATTATA ATTGTACATT ACGTACTGCA ATAAATACTC | BssHII       |
| 2031 | ATGGGTTTT ATGATTAGAG TCCCGAATT ATACATTAA TAGCGATAG AAAACAAAT ATAGCGCGCA<br>TACCCAAAAA TACTAACTC AGGGCGTTAA TATGTAATT ATGCGCTATC TTTTGTTTA TATCGCGCGT         | XbaI         |
| 2101 | AACTAGGATA AATTATCGCG CGCGGTGTCA TCTATGTAC TAGATCGATA AGCTTCTAGA GCGGCCGGTG<br>TTGATCCTAT TTAAATAGCGC GCGCCACAGT AGATAACATG ATCTAGCTAT TCGAAGATCT CGCCGGCCAC | BssHII       |
| 2171 | GAGCTCCAAT TCGCCCTATA GTGAGTCGTA TTACGGCGCG TCAACTGGCC TCGTTTTACA ACGTCTGTGAC<br>CTCGAGGTAA AGCGGGATAT CACTCAGCAT AATGCGCGG AGTGAACGGC AGCAAAATGT TGCAGCACTG | Clai HindIII |
| 2241 | TGGGAAACCC CTGGCGTTAC CCAACTTAAT CGCCCTTGAG CACATCCCCC TTTCGCCAGC TGGCGTAATA<br>ACCCTTTTG GACCGCAATG GGTGAAATT GCGGAACGTC GTGTAGGGGG AAAGGGTGG ACCGGATTAT    | BssHII       |
| 2311 | GCGGAAGAGGC CCGCACCGAT CGCCCTTCCC AACAGTTGGG CAGCCTGAAT GGGAAATGGG ACGGCCCTG<br>CGCTTCTCCG GCGTGGCTA GCGGAAGGG TTGTCAACGC GTGGACTTA CGCTTACCC TGCGGGGAC      | BssHII       |
| 2381 | TAGGGGGCA TAAAGCGGG CGGGTGTGGT GGTACGGC AGGGTGAACCG CTACACTTGC CAGGCCCTA<br>ATCGCCCGT AATTGGGCC GCCCACACCA CCAATGGCG TCGCACTGGC GATGTGAACG GTCGGGGAT         | BssHII       |
| 2451 | GGCCCCGGTC CTTTCGCTTT CTTCCCTTC TTTCTCGCCA CGTTCGCCCG CTTTCCCCGT CAAGGTCTAA<br>CGGGGGGAG GAAAGGGAAA GAAGGAAGG AAAGAGGGT GCAAGGGCC GAAAGGGCA GTTCGAGGATT      | BssHII       |

FIG.-42E

2521 ATCGGGGCT CCCTTTAGGG TCCCGATTAA GTGCTTTACG GCACCTCGAC CCCAAAAAC TTGATTAGGG  
TAGCCCCGA GGGAAATCCC AAGGCTAAAT CACGAAATGCG CGTGGAGCTG GGGTTTTG AACTAATCCC

2591 TGATGGTTCA CGTAGTGGGC CATGCCCTG ATAGACGGTT TTTCGCCCC TTGACGTTGGA GTCCACGTTTC  
ACTACCAAGT GCATCACCCG GTAGGGAC TATCTGCCAA AAAGGGAA ACTGCAACCT CAGGTGCAAG

2661 TTTAATAGTG GACTCTTGT CCAAACCTGGA ACAACACTCA ACCCTATCTC GGCTTATTCT TTGATTAT  
AAATTATCAC CTGAGAACAA GGTGTGACCT TGGTGTGAGT CCAGATAAGA AAACAAATA

2731 AAGGGATTIT GCCGATTTCG GCCTATTGGT TAAAAATGA GTGTGATTAA CAAAATTA ACGGAAATT  
TTCCCTAAA CGGCTAAAGC CGGATAACCA ATTTCCTTACT CGACTAAATT GTTTTAAAT TGCGCTTAAA

2801 TAACAAAATA TAAACGCTTA CAATTAGGT GGGCACTTTTC GGGAAATGT GCGGGAAACC CCTATTGTT  
ATTTGTTTAT AATTGCGAAT GTAAATCCA CGTGAAAAG CCCCTTACA CGGCCCTTGG GGATAAACAA

2871 TATTTTCTA AATACATTCA AATATGTATC TTATACATAG CGCTCATGAG ACAATAACCC TGATAATGCT  
ATAAAAAGAT TTATGTAAGT TTATACATAG GCGAGTACTC TGTTATTGGG ACTATTACG TTCAATAATA  
AACTTTTTC TTCTCATACT CATAAGTTGT AAAGGCACAG CGGGAAATAAG GGAAAAACG AAGTTATTAT

2941 TTGAAAAGG AAGAGTATGA GTATTCAACA TTCCCGTGT GCCTTATTC CCTTTTTGC GGCATTG  
AACTTTTTC TTCTCATACT CATAAGTTGT AAAGGCACAG CGGGAAATAAG GGAAAAACG CCGTAAACG

3011 CTTCCCTGTT TTGCTCACCC AGAAACGCTG GTGAAAGTAA AGATGCTGA AGATCAGTT  
GAAGGACAAA AACGAGTGGG TCTTGGAC CACTTCTATT TTCTACGACT TCTAGTCAAC CCACGTTG  
ACCCAAATGTA GCTTGACCTA GAGTTGTCGC CATTCTAGGA ACTCTCAAAG GCGGGCTTC TTGCAAAGG

3081 TGGGTACAT CGAACTGGAT CTCAACAGCG GTAAGATCCT TGAGAGTTT CGCCCCGAAG AACGTTTCC  
TTACTACTCG TGAAAATTTC AAGACGATAAC ACCGCGCCAT AATAGGGCAT AACTGTCAAAC CGTTCTCGTT

3151 AATGATGAGC ACTTTAAG TTCTGCTATG TGGCGGGTA TTATCCCGTA TTGACGCCGG GCAAGGGCAA  
TTACTACTCG TGAAAATTTC AAGAGTCTTA CTGAACCAAC TCATGAGTGG TCAGTGTCTT TTGCTAGAAT

3221 CTCGGTGCCT GCATACACTA TTCTCAGAAAT GACTTGGTTG AGTACTCACC AGTCACAGAA AAGCATCTTA  
GAGCCAGGG CGTATGTTGAT AAGAGTCTTA CTGAACCAAC TCATGAGTGG TCAGTGTCTT TTGCTAGAAT

3291 CGGATGGCAT GACAGTAAGA GAATTATGCA GTGCTGCCAT ACCATGAGT GATAACACTG CGGCCAACTT  
GCCTACCGTA CTGTCATCT CTTAATACGT CACGACGGTA TTGGTACTCA CTATTGTCAC GCGGGTTGAA

FIG.-42F

3361 ACTTCTGACA ACGATCGGAG GACCGAAGGGA GCTAACCGCT TTTTGCACA ACATGGGGA TCATGTTA  
TGAAAGACTGT TGCTAGCCTC CGGGCTTCCT CGATTGGCA AAAAACGTGT TGTACCCCT AGTACATTGA

3431 CGCCTTGATC GTTGGGAACC GGAGCTGAAT GAAGCCATAC CAAAGACGA GCGTGACACC ACGATGCCCTG  
GCGGAACTAG CAAACCTTGG CCTCGACTTA CTTCGGTATG GTTGTGTGT CGCAGCTGTGG TGCTACGGAC

3501 TAGGAATGGC AACAAACGTTG CGCAAACAT TAACCTGGGA ACTACTTACT CTAGCTTCCC GGCAACAATT  
ATCGTTACCG TTGTTGCAAC GCGTTTGATA ATTGACCGCT TGATGAATGA GATCGAAGGG CGGTGTTAA

3571 AATAGACTGG ATGGAGGGGG ATAAAGTTGC AGGACCACTT CTGGCCTCGG CCCTTCCGG TGCTGGTT  
TTATCTGACCC TACCTCCGCC TATTCAACG TCCTGGTGAAC GACGGCAGGCC GGGAAAGGG ACCGACCAAA

3641 ATTGCTGATA AATCTGGAGC CGGTGAGCGT GGGTCTCGG GTATCATTGC AGCAGCTGGG CCAGATGGTA  
TAAAGACTAT TTAGACCTCG GCCACTCGCA CCCAGAGGGC CATACTAACG TCGTGACCCC GGTCTACCAT

3711 AGCCCTCCCG TATCGTAGTT ATCTACACGA CGGGGAGTCA GGCAACTATG GATGAACGAA ATAGACAGAT  
TGGGGGGC ATAGCATCAA TAGATGTGCT GCCCCTCACT CGGTGATAC CTACTTGCTT TATCTGTCTA

3781 CGCTGAGATA GGTGCCTCAC TGATTAAGCA TTGGTAACCTG TCAAGACCAAG TTTACTCATA TATACCTTAG  
GCGACTCTAT CCACGGAGTG ACTAATTCTGT ACCATTGAC AGTCTGGTTC AAATGAGTAT ATATGAATC

3851 ATTGATTAA AACTCATTT TAAATTAAA AGGATCTAGG TGAAAGATCTCT TTTTGATAAT CTCATGACCA  
TAACTAAATT TTGAAAGTAA AATTAAATT TCCTAGATCC ACTTCTAGGA AAACATTAA GAGTACTGGT

3921 AAATCCCTTA ACGTGAGTTT TCGTTCCACT GAGCGTCAGA CCCCGTAGAA AAGATCAAAG GATCTCTTG  
TTAGGGAAAT TGCACTCAA AGCAAGGTGA CTCGCAGTCT GGGGCATCTT TTCTAGTTTC CTAGAAGAAC

3991 AGATCCTTT TTTCTGCGCG TAATCTGCTG CTTGCAAACA AAAAACACAC CGCTACCGC GGTGGTTGT  
TCTAGGAAA AAAGACGGGC ATTAGACGAC GAACGTTGT TTTTTGGTG GCGATGGTGC CCACCAAACA

4061 TTGCCGGATC AAGAGCTACC AACTCTTTT CGAAGGGTA CTGGCTTCAG CAGAGGCAG ATACCAAATA  
AACGGCCTAG TTCTCGATGG TTGAGAAAA GGCTTCCATT GACCGAAAGTC GTCTCGCGTC TATGGTTAT

4131 CTGTCCTTCT AGTGTAGGCC TAGTTAGGCC ACCACTCAA GAACCTCTGTA GCACCGCCTA CATACTCGC  
GACAGGAAGGA TCACATCGGC ATCAATCCGG TGGTGAAGTT CTGAGACAT CGTGGGGAT GTATGGCG

FIG.-42G

4201 TCTGCTTAATC CTGTTACCAAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGACTCAAGA  
AGACGATTAG GACAATGGTC ACCGACGACG GTCACCGCTA TTCAAGCACAG AATGGCCAA CCTGAGTTCT  
4271 CGATAGTTAC CGGATAAGGC GCAGGGTCTG GGTGAACCG GGGTTCGTG CACACGCC AGCTGGAGC  
GCTATCAATG GCCTATTCCG CGTCGCCAGC CCGACTTGCC CCCAAGCAG GTGTTGCGG TCGAACCTCG  
4341 GAACGACCTA CACCGAACTG AGATAACCTAC AGCGTGAAGCT ATGAGAAAGC GCCACGCTTC CGAACGGAG  
CTTGCTGGAT GTGGCTTGAC TCTATGGATG TCGCACTCGA TACTCTTCG CGTGGGAAG GGCTTCCCTC  
4411 AAAGGGGAC AGGTATCCGG TAAGGGCAG GGTGGAAACA GGAGAGCGCA CGAGGGAGT TCCAGGGGA  
TTTCCGGCTG TCCATAGGCC ATTGCGCTC CGAGCCTGT CCTCTCGCGT GCTCCCTCGA AGGTCCCCCT  
4481 AACGGCCTGGT ATCTTTATAG TCCTGTGGGG TTTCGCCACC TCTGACTTGA GCGTCGATT TTGTTGATGCT  
TTGGGACCA TAGAAATATC AGGACAGCCC AAAGCAGCCC AGACTGAAC TCGAGCTAA AACACTACGA  
4551 CGTCAGGGG GCGGAGCCTA TGGAAAACG CGCAACCGC GGCTTTTA CGGTTCTGG CCTTTGCTG  
GCAGTCCCCC CGCCTCGGAT ACCTTTTGC GTCGGTGGC CGGAAAAAT CGCAGCTAA AACACTACGA  
4621 GCCTTTGCT CACATGTTCT TTCCCTGCGTT ATCCCCTGAT TCTGTTGGATA ACCGATTAC CGCCTTGAG  
CGGAAAACGA GTGTACAAGA AAGGACGCAA TAGGGGACTA AGACACCTAT TGGCATAATG GCGGAAACTC  
4691 TGAGCTGATA CCGCTCGCCG CAGCCGAACG ACCGAGCGCA CGGAGTCAGT GAGGGAGGAA GCGGAAGAGC  
ACTCGACTAT GGCGAGGGC GTGGCTTGC TGGCTCGGT CGCTCAGTC CCTGGCTCCTT CGCCTCTCG  
4761 GCCCAATAAG CAAACCGCCT CTCCCCGCCG GTGGCCGAT TCATTAATGC AGCTGGCAGG ACAGGGTTCC  
GGGTTATGC GTTGGCGGA GAGGGCGCG CAACCGCTA AGTAATTACG TCGACCGTGC TGTCCAAAGG  
4831 CGACTGGAAA GCGGGCACTG AGCGCAACGC AATTAATGTG AGTTAGCTCA CTCATTAGGC ACCCCAGGCT  
GCTGACCTT CGCCGTCAC TCGCGTTGC TAAATTACAC TCAATCGAGT GAGTAATCCG TGGGGTCCGA  
4901 TTACACTTAA TGCTTCCGGC TCGTATGTTG TGAGGGATA ACAATTTCAC ACAGGAAACA  
AATGTGAAAT ACGAAGGGCG AGCATACAA ACACCTTAAC ACTCGCCTAT TGTTAAAGTG TGTCCTTGT

FIG.-42H

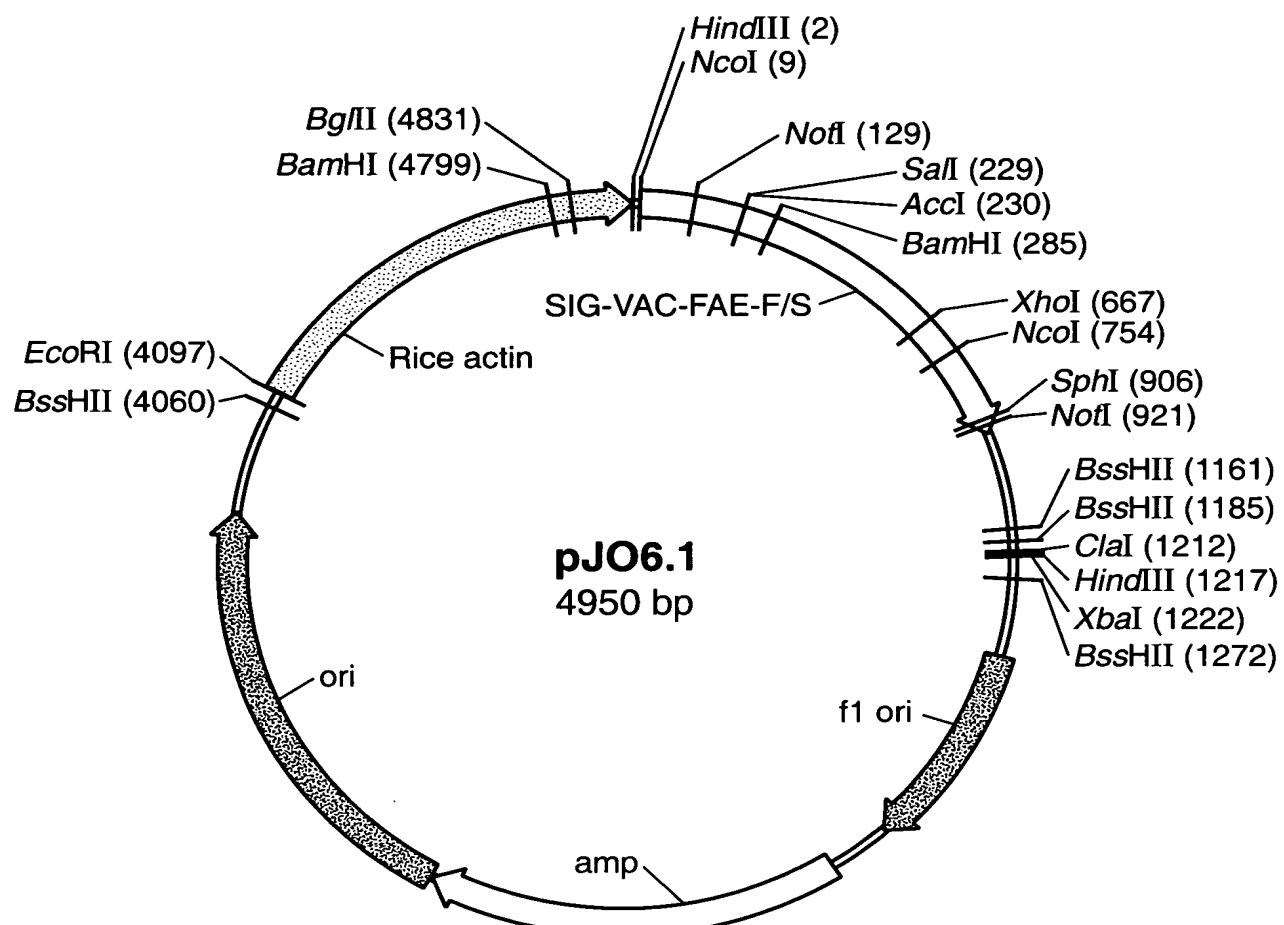
4971 GCTATGACCA TGATTAACGCC AAGCGGCCAA TTAACCCCTCA CTAAGGGAA CAAAGCTGG GTAC  
CGATACTGGT ACTAATGGGG TTTCGGCGTT ATTGGGAGT GATTTCGCCCTT GTTTTCGACCC CATG

BssHII ~~~~~  
KpnI ~~~~~

NcoI

**FIG.\_421**

100 / 154



**FIG.\_43A**

HindIII NcoI

|     |                                                 |                            |                          |                           |                           |                           |
|-----|-------------------------------------------------|----------------------------|--------------------------|---------------------------|---------------------------|---------------------------|
| 1   | AAGCTTACCA TGGCCCCACGC<br>TTCGAATGGT ACCGGTGGG  | CCGGGTCCCTC<br>GGGGCAGGGAG | CTCCTGGCGC<br>AGGGCACCGG | TGCGCGT<br>GAGGACCGA      | GGCCACGGCC<br>CCGGTGGCGG  | GCCGTGCGCG<br>GGGCAGGGC   |
|     |                                                 |                            |                          |                           | Not I                     |                           |
| 71  | TCGCCCTCCTC CTCCCTCCCTC<br>AGGGAGGAG GAGGAGGAAG | GGCGACTCCA<br>CGGTGAGGGT   | ACCCGATCCG<br>TGGGCTAGGC | GCCCCGTACCC<br>CGGGCAGTGG | GACCGGGCGG<br>CTGGCGGCC   | CCGCCCTCCAC<br>GGGGAGGGT  |
| 141 | GCAGGGCCTC TCCGAAGGACC<br>CGTCCCGTAG            | TCTACAGGCC<br>AGGCTCTGG    | TTTACAGCG<br>AAATCAGCTT  | ATGGCCACTA<br>TACCGGTAT   | TCTCCCAAGC<br>AGAGGGTTCG  | TGCCCTACGCC<br>ACGGATGCGG |
|     |                                                 |                            |                          |                           | Sall                      |                           |
|     |                                                 |                            |                          |                           | AccI                      |                           |
| 211 | GACCTGGC<br>CTGGACACGT                          | ACATTCCGTC<br>TGTAAGGCAG   | GACTATTATC<br>CTGATAATAG | AAGGGAGAGA<br>TTCCCTCTCT  | AAATTACAA<br>TTAAATGTT    | TTCTCAAATC<br>AAGAGTTGA   |
|     |                                                 |                            |                          |                           | BamHI                     |                           |
| 281 | GATGGATCCT<br>CTACCTAGGA                        | CGCGGACGAC<br>GGCGCTGCTG   | AGCAGCAAG<br>TGTGCGTTTC  | AAATAATCAC<br>TTTATTAGTG  | CGTCTTCCGT<br>GCAGAAGGCA  | GGCACTGGTA<br>CCGTGACCAT  |
| 351 | TCTACAACTC<br>AGATGGTGGAG                       | GATACTAACT<br>CTATGATGTA   | ACACCCTCAC<br>TGTGGAGTG  | GCCTTTCGAC<br>CGGAAAGCTG  | ACCCCTACAC<br>TGGGATGGTG  | AATGCAACGG<br>TTACGTTGCC  |
| 421 | CACGGTGGAT<br>GTGCCACCTA                        | ATTATAATTGG<br>TAATAAACCT  | ATGGGTCTCC<br>TACCCAGAGG | GTCCAGGACC<br>CAGGTCTGG   | AAGTCGAGTC<br>TTCAAGCTCAG | GCTTGTCAA<br>CGAACAGTTT   |
| 491 | GCCAGTATCC<br>CGGTCAATAGG                       | GGACTACGGC<br>CCTGATGGC    | CTGACCGTGA<br>GACTGGCACT | CGGGCCACKC<br>GGCGGGTGMG  | CCTCGGGGCC<br>GGAGGCGGG   | TCCCTGGCGG<br>AGGGACCGCC  |
| 561 | CGCCCAAGCTG<br>CGGGGTGAC                        | TCTGGGACAT<br>AGACGCTGAC   | ACGACAACAT<br>TGCTGTTGTA | CCGGCTGTAC<br>GGGGACATG   | ACCTTCGGCG<br>TGGAAAGCCG  | AACCGGGCAG<br>TTGGCGTC    |

FIG.-43B

XbaI

631 GCCTTCGGT CGTACATGAA CGATGCCCTTC CAAGCCTCGA GCCCAGATA GACGGAGTAT TTCCGGGTCA  
CGGAAGCGCA GCATGTACTT GCTACGGAAG GTTCTATG CGGGTCTATG CTGCGTCATA AAGGCCAGT

NcoI

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCCGGGGA GCAGGGGTAC CCCATGGCG GTGTAGAGTA  
GAGTAGGTT GCTGCCGTAG GGTTGGACGG GGGCCACCT CGTCCCCATG CGGGTACCGC CACATCTCAT

771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCACTGGG ATGAAAGTGC GTGCTGTGAG  
GACCTCGCAA CTAGGAATGT CGGGGGTCT GTGTAAACAG ACGTGACCCCC TACTTCACGT CACGACACTC

SphI

841 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGGGC GCATGCACCT  
CGGGTCCCGC CTGTCCCCACA CTTATTACGC GTGTGCTGAA TAAAACCTA CTGCTCGCCG CGTACGTGGA

Not I

911 GGCCGGTTCGC GGCGGGGAA ACCACTGAAG GATGAGCTGT AAAGAAGGCAG ATCGTTCAA CATTGGCAA  
CGGGCCAGCG CGGGCGCTT TGGTGACTTC CTACTCGACA TTCTTCGTC TAGCAAAGTT GTAAACCGTT

981 TAAAGTTCT TAAGATTGAA TCCCTGTTGCC GGTCTTGCGA TGATTATCAT ATAATTTCG TTGAATTACG  
ATTTCAGAA ATTCTAACTT AGGACAACGG CCAGAACGCT ACTAATAGTA TATTAAGAC AACTTAATGC

1051 TTAAGGATGT AATAATTAC ATGTAATGCA TGACGGTTATT TATGAGATGG GTTTTTATGA TTAGAGTCCC  
AATTCTGACAA TTATTAATG TACATTACGT ACTGCAATAA ATACTTACCA AAAAATACT AATCTCAGGG

BssHII

1121 GCAATTATAC ATTAATACG CGATAGAAA CAAATAATAG CGCGCAAAACT AGGATAAAATT ATCGCGCGCG  
CGTTAAATATG TAAATTATGC GCTATCTTT GTTTTATATC GGGCGTTTGA TCCTTATTAA TAGCGCGCGC

FIG.-43C

|      |                                                                                             | XbaI                                                                   | Clal HindIII |
|------|---------------------------------------------------------------------------------------------|------------------------------------------------------------------------|--------------|
| 1191 | GTGTCATCTA TGTTACTAGA TCGATAAGCT TCTAGAGCGG<br>CACAGTAGAT ACAAATGATCT AGCTATCGA AGATCTCGCC  | CCGGTGGAGC TCCAATT CGC CCTATAGTGA<br>GGCCACCTCG AGGTAAAGCG GGATATCACT  |              |
|      |                                                                                             |                                                                        | BssHII       |
| 1261 | GTCGTATTAC GCGGGCTCAC TGGCCGTCGT TTTACAAACGT<br>CAGGATAATG CGGGCGAGTG ACCGGCAGCA AAATGTTGCA | CGTGACTGGG AAAACCCCTGG CGTTACCCAA<br>GCACGTGACCC TTTGGGACCC GCAATGGGTT |              |
| 1331 | CTTAATCGCC TTGCGAGCAC TCCCCCTTTC GCCAGCTGGC<br>GAATTAGCGG AACGTCGTGT AGGGGGAAAG CGGTCGACCG  | GTAATAGCGA AGAGGCCCGC ACCGATCGCC<br>CATTATGCT TCTCCGGGCG TGGCTAGGG     |              |
| 1401 | CTTCCCAACA GTTGGCGAGC CTGAATGGCG AATGGGACGC<br>GAAGGGTTGT CAACGGCTCG GACTTACCGC TTACCTCGG   | GCCCTGTAGC GGGGACATAA GCGGGCGGG<br>CGGGACATCG CGCGTAATT CGGCCGCGCC     |              |
| 1471 | TGTGGTGGTT ACGGCGAGCG TGACCCGCTAC ACTTGGCAGC<br>ACACCAACCA TGCGCGTGC ACTGGCGATG TGAACGGTGC  | GCCCTAGCGC CGGCTCCCTT CGCTTTCTTC<br>CGGGATCGCG GGCAGGGAA GCGAAAGAAG    |              |
| 1541 | CCTTCCTTTC TCGCCACGTT CGCCGGCTT CCCCCGTCAAG<br>GGAAGGAAAG AGGGGTGCAA GCGGCCGAAA GGGCAGTTC   | CTCTAAATCG GGGGCTCCCT TTAGGGTTCC<br>GAGATTAGC CCCGAGGGAA AATCCCAAGG    |              |
| 1611 | GATTAGTGC TTACGGCAC CTCGACCCCA AAAAAGCTTA<br>CTAAATCACG AAATGCCGTG GAGCTGGGT TTTTGAACT      | TTAGGGTGTAT GTTTCACGTA GTGGGCCATC<br>AATCCCAACTA CCAAGTGCAT CACCCGGTAG |              |
| 1681 | GCCCTGTAG ACGGTTTC GCCTTTGAC GTTGGAGTCC<br>CGGGACTATC TGCCAAAAG CGGGAAACTG CAACCTCAGG       | ACGTTCTTTA ATAGTGGACT CTTGTTCCAA<br>TGCAAGAAAT TATCACCTGA GAAACAAGGT   |              |
| 1751 | ACTGGAAACAA CACTCAACCC TATCTGGTC TATTCTTTG<br>TGACCTGTGT GTGAGTGGG ATAAGCCAG ATAAGAAAAC     | ATTATAAGG GATTTTGGCG ATTTCGGCCCT<br>TAAATATTCC CAAAACGGC TAAAGCGGA     |              |
| 1821 | ATTGGTTAAA AAATGAGGTG ATTAAACAA AATTAAACGC<br>TAACCAATT TTACTCGAC TAAATGTT TAAATTGCG        | GAATTAAAC GAAATATAAA CGCTTACAAAT<br>CTTAAATTG TTTATAATT GCGAATGTTA     |              |

1891 TAGGGTGGCA CTTTCGGGG AAATGTGGC GGAACCCCTA TTTGTTTATT TTTCTAAATA CATTCAAAATA  
AATCCACCGT GAAAAGCCCC TTACACCGCG CCTTGGGGAT AAACAAATA AAAGATTAT GAAAGTTTAT  
+  
1961 TGTATCCGCT CATGAGACAA TAACCCTGAT AAATGCTTCATA ATAATATTGA AAAAGGAAGA GTATGAGTAT  
ACATAGGGCA GTACTCTGTT ATGGGACTA TTACGAAGT TATTATAACT TTTCCCTTCT CATACTCATA  
2031 TCAACATTTC CGTGTGGCCC TTATTCCCTT TTTGCGGCA ATAATATTGA AAAAGGAAGA GTATGAGTAT  
AGTTGAAAG GCACAGGGG AATAAGGGAA AAAACGGCGT AAAACGGAA GACAAAAACG AGTGGGTCTT  
2101 ACGCTGGTGA AAGTAAGAAGA TGCTGAAGAT CAGTTGGGTG CACGAGTGGG TTACATCGAA CTTGGATCTCA  
TGCACCACT TTCAATTCT ACGACTTCATA GTCAACCCAC GTGCTCACCC AATGTAGCTT GACCTAGAGT  
2171 ACAGGGTAA GATCCTTGAG AGTTTTCGGC CCGAAAGAACG TTTTCCAATG ATGAGCACTT TAAAGGTCT  
TGTCGCCATT CTAGGAACCTC TCAAAGGG GGCTTCTTGC AAAAGGTTAC TACTCGTGA AATTCAAGA  
2241 GCTATGTGGC GCGGTATTAT CCGTATGTA CGCCATAACT CGCCGGCAA GAGCAACTCG GTCGGCGCAT ACACATCT  
CGATACACCG CGCCATAATA GGGCATAACT GCGGCCCGTT CTCGTTGAGC CAGGGCGTA TGTGATAAGA  
2311 CAGAATGACT TGGTTGAGTA CTCACCACTC ACAGAAAAGC ATCTTACGGA TGGCATGACA GTAAGAGAAT  
GTCTTACTGA ACCAACTCAT GAGTGGTCAG TGTTTTTCC TAGATGCGCT ACCGTACTGT CATTCTCTTA  
2381 TATGCACTGC TGCCATAACC ATGAGTGTATA ACACGTGGC CAACTTACTT CTGACAACGA TCGGAGGACC  
ATACGTCACTG ACGGTATTGG TACTCACTAT TGTAATGAA GACTGTTGCT AGCCTCCTGG  
2451 GAAGGGAGCTA ACCGGCTTTTG TGCAACACAT GGGGGATCAT GTAAACTCGCC TTGATCGTTG GAAACCGGAG  
CTTCCTCGAT TGGGAAAAA ACGTGTTGTA CCCCTAGTA CATTGAGCGG AACTAGAAC CCTTGGCCTC  
2521 CTGAATGAAG CCATACCAAA CGACGAGGGT GACACCACGA TGCCTGTAGC AATGGCAACA ACGTTGCGCA  
GACTTACTTC GGTATGGTTT GCTGCTCGCA CTGGGTGCT ACGGACATCG TTACCGTTGT TGCAACCGGT  
2591 AACTATTAAC TGGCGAACTA CTTACTCTAG CTTCCGGCA ACAATTAATA GACTGGATGG AGGGGATAAA  
TTGATAATTG ACCGCTTGAT GAATGAGATC GAAGGGCGT TGTTAATTAT CTGACCTACC TCCGGCTATT  
2661 AGTTGGAGGA CCACCTCTGC GCTCGGCCCT TCCGGCTGGC TGGTTTATTG CTGATAATC TGGAGCCGGT  
TCAACGTCCCT GGTGAAGACG CGAGCCGGGA AGGGCGACCG ACCAAATAAC GACTATTAG ACCTCGGGCCA

2731 GAGCGTGGGT CTCGGGGT CATTGCAGCA CTGGGGCCAG ATGGTAAGCC CTCCCGGTATC GTAGTTATCT  
CTCGCACCCA GAGGCCATA GTAACTGCTGT GACCCGGTC TACCATTCGG GAGGGCATAG CATCAATAGA

2801 ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAATAG ACAGATCGCT GAGATAGGTG CCTCACTGAT  
TGTGCTGCC CTCAGTCGGT TGATACCTAC TTGCTTAGCCTA CTCTATCCAC GGAGTGACTA

2871 TAAGCATTGG TAACTGTCAG ACCAAGTTA CTCATATATA CTTAGATTG ATTAAACT TCATTTTAA  
ATTCGTTAACC ATTGACAGTC TGTTCAAAT GAGTATATACTAAC TAAATTGTA AGTAAAAATT

2941 TTTAAAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA TGACCCAAT CCCTTAACGT GAGTTTTCGT  
AAATTTCCT AGATCCACTT CTAGGAAAAA CTATTAGGT ACTGGTTTA GGGAAATTGCA CTCAAAGGCA

3011 TCCACTGAGC GTCAGACCCC GTAGAAAGA TCAAAGGATC TTCTTGAGAT CTTTTTTTC TGCGCGTAAT  
AGGTGACTCG CAGTCTGGG CATTTTCT AGTTTCCTAG AAGAACTCTA GGAAAAAAG ACGGGCATTAA

3081 CTGCTGCTTG CAAACAAAA ACCACCGCT ACCAGGGTG GTTGTGTTGC CGGATCAAGA GCTACCAAAT  
GACGAGAAC GTTGTGTTT TTGGTGGCGA TGGTCGCCAC CAAACAAACG GCCTAGTTCT CGATGGTTGA

3151 CTTTTCCGA AGGTAACTGG CTTCAAGCAGA GGCAGAGATAC CAAATACTGT CCTTCTAGTG TAGCGCTAGT  
AAAAAGGCT TCCATTGACC GAAGTCTGT CGCGTCTATG GTTTATGACA GGAAGATCAC ATGGTCACCG

3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCCTACATA CCTCGCTCTG CTAATCTGT TACCACTGGC  
ATCCGGGGT GAAGTCTTG AGACATCGTG GCGGATGTTAT GGAGGGAGAC GATTAGGACA ATGGTCACCG

3291 TGCTGCCAGT GGCAGATAAGT CGTGTCTTAC CGGGTTGGAC TCAAGACGAT AGTTACGGAA TAAGGGCAG  
ACGAGGTCA CCGCTATTCA GCACAGAAATG GCCCAACCTG AGTTCTGCTA TCAATGGCCT ATTCCGGTC

3361 CGGTGGGT GAACGGGGG TTCTGTGCACA CAGGCCAGCT TGGAGCGAAC GACCTACACC GAACTGAGAT  
GCCAGGGCA CTGGCCCCC AAGGACGTGT GTCGGGTCGA ACCTCGCTTG CTGGATGTTG CTTGACTCTA

3431 ACCTACAGGG TGAGCTATGA GAAAGGGCCA CGCTTCCCGA AGGGAGAAAG GGGGACAGGT ATCCGGTAAG  
TGGATGTCGC ACTCGATACT CTTTCGGGT GCGAAGGGCT TCCCTTTTC CGCCTGTCCA TAGGCCATTTC

3501 CGGCAGGGC GGAACAGGGAG AGGGCACGGAG GGAGCTTCCA GGGGAAACG CCTGGTATCT TTATAGTCT  
GCCGTCCAG CCTTGTCCTC TGCGTGTCTC CCTCGAAGGT CCCCTTTGC GGACCATAGA ATATCAGGA

3571 GTCGGGTTTC GCCACCTCTG ACTTGAGCGT CGATTTTGT GATGCTCGTC AGGGGGGG AGCCTATGGA  
CAGCCAAAG CGGGAGAC TGAACTCGCA GCTAAAAACA CTACGAGCAG TCCCCCGCC TCGGATAACCT

3641 AAAACGCCAG CAACGGGCC TTGTTACGGT TCCCTGGCCT TTGCTGGCCT TTGCTCACA TTGCTCTTCC  
TTTGGGTC GTTGGCGG AAAAATGCGA AGGACCGGAA AACGACCGGA AAACGAGTGT ACAAGAAAGG

3711 TGGTGTATCC CCTGATTCTG TGGATAACCG TATTAACGCC TTGAGTGAG CTGATAACCGC TGCCGCAGC  
ACGCAATAGG GGACTAAGAC ACTTATGGC ATAATGGCG AAACTCACTC GACTATGGG AGCGGGCTCG

3781 CGAACGACCG AGGGCAGCGA GTCAAGTGAGC GAGGAAGCGG AAGAGGCC AATAACGAAA CGGCCTCTCC  
GCTTGGTGGC TCGGTGCT CAGTCACCTG CTCCTCGCC TTCTCGGGG TTATGGTTT GGGGGAGAGG

3851 CCGCGGTTG GCGGATTCA TATGCGACT GGCACCGACAG GTTCCCGAC TGGAAAGGG GCAGGTGAGCG  
GGCGGCAAC CGGCTAAGTA ATTACGTGCA CCGTGCTGTC CAAAGGGCTG ACCTTTCGCC CGTCACTCGC

3921 CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTAC ACTTTATGCT TCCGGCTCGT  
GTTGGTTAA TTACACTCAA TCGAGTGAGT AATCCGTTGGG GTCCGAAATG TGAATATACGA AGGCCGAGCA

~~~~~ BssHII ~~~~~

3991 ATGTTGTGAG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC
TACAAACAC CTTAACACTC GCCTATTGTT AAAGTGTGTC CTTTGTGAT ACTGGTACTA ATGCGGGTTCG

~~~~~ BssHII ~~~~~

4061 GCGCAATTAA CCCTCACTAA AGGGAACAAA AGCTGGAAATT CCACAAATGAA CAATAATAAG ATTAAAATAG  
CGCGTTAATT GGGAGTGATT TCCCTTGTGTT TCGACCTTAA GGTTGTTACTT GTTATTATTC TAATTTTATC

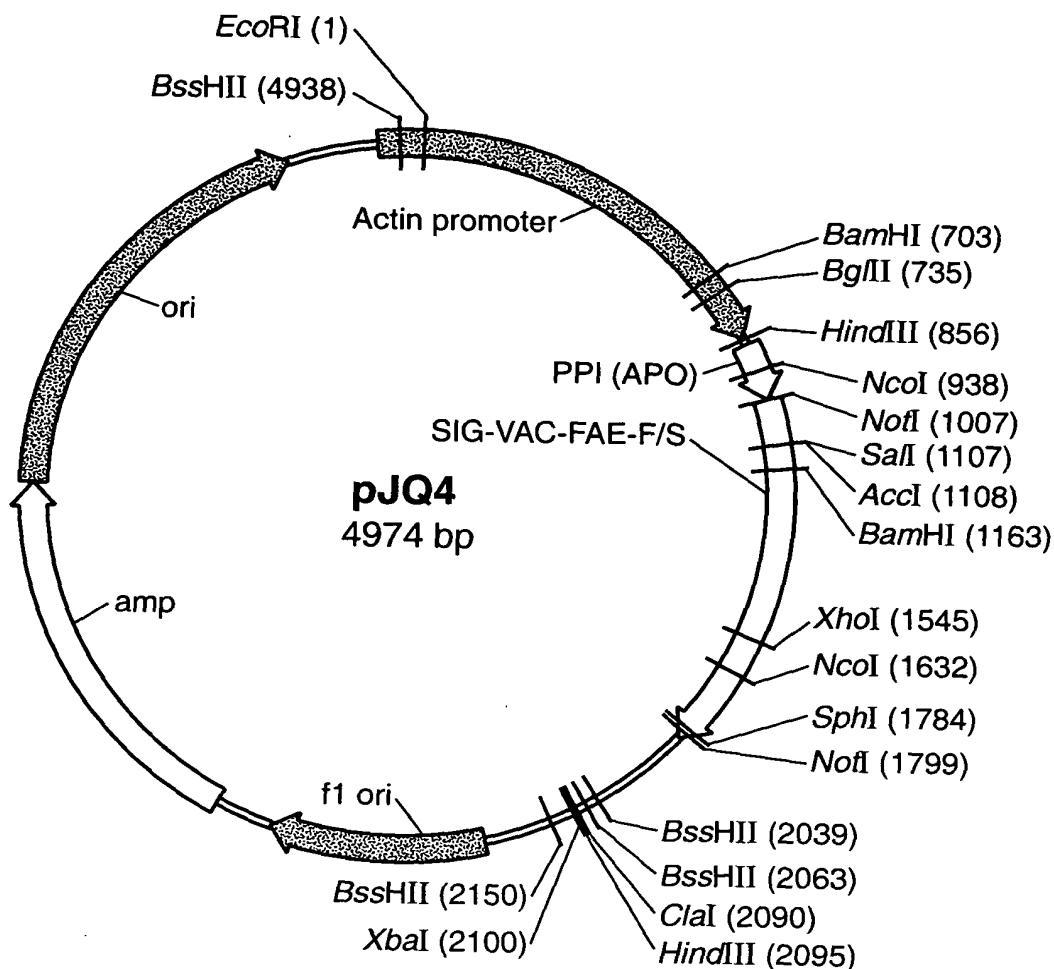
4131 CTTGCCCGG TTGCAGCGAT GGGTATTTT TCTAGTAAA TAAAAGATA ACTTAGACTC AAAACATTAA  
GAACGGGGC AACGTCGCTA CCATAAAAA AGATCATT ATTTCCTATT TGAATCTGAG TTTTGTAAAT

4201 CAAACAAAC CCCTAAAGTC CTAAGGCCA AAGTGCTATG CACGATCCAT AGCAAGGCCA GCCCAACCCA  
GTTTGTGTTG GGGATTTCAG GATTTCGGGT TTCAACGATAAC GTGCTAGGTA TCGTTGGGT CGGGTTGGGT

FIG.\_43G

|      |                                                                                                                                                                                           |
|------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 4271 | ACCCAAACCCA ACCCACCCCA GTGCAGCCAA CTGGCAAATA GTCTCCACCC CCGGCACCAT CACCGTGA<br>TGGGTGGGT TGGGTGGGT CACGTCGGTT GACCGTTTAT CAGAGGTGGG GGCCGTGATA GTGGCACTCA                                 |
| 4341 | TGTCCGCACC ACCGCACGTC TCGCAGCCAA AAAAAGAAAA AGAAAGAAAA AAAAGAAAAA GAAAACAGC<br>ACAGGGTGG TGGCGTGCAG AGCGTCGGTT TTTTTTTT TCTTTCTTTT CTTTTTTGTCG                                            |
| 4411 | AGGTGGTCC GGGTGTGGG GGGCGGAAA GCGAGGGAGGA TCGCGAGCAG CGACGGGCC CGGCCCTCCC<br>TCCACCCAGG CCCAGCAGG CGCGCCTTT CGCTCCTCCT AGCGCTCGTC GCTGCTCCGG GCCGGGAGGG                                   |
| 4481 | TCCGGTTCCA AAGAAACGCC CCCATCGCC ACTATATACA TACCCCCC TCTCCTCCCA TCCCCAAC<br>AGGCAGAGGT TTCTTGGGG GGGTAGGG TGATATATGT ATGGGGGG AGAGGAGGGT AGGGGGTTG                                         |
| 4551 | CCTACCACCA CCACCAAC CACCTCCTCC CCCCTCGCTG CGGGAGCAG AGCTCCCTCCC CCCTCCCCCT<br>GGATGGTGGT GGTGGTGGT GTGGAGGG GGGAGGGAC GGCCTGCTGC TCGAGGAGGG GGGAGGGGGA                                    |
| 4621 | CGGCCGCCGC CGGTAACCCAC CCCGCCCTC TCCTCTTTCT TTCTCCGTTT TTTTTTCGTT CTCGGTCTCG<br>GGGGGGGGG GCCATTGGTG GGGGGGGAG AGGAGAAAGA AAGAGGAAA AAAAAGCA GAGCCAGAGC                                   |
| 4691 | ATCTTGGCC TTGGTAGTT GGTTGGCGA GAGGGCTTC GTGCCCCAGA TCGGGGGCG GAGGGGGGG<br>TAGAACCGG AACCATAAA CCACCCGCT CTCGCCGAAG CAGGGGTCT AGCCACGGGC CCTCCCCGCC                                        |
|      | BamHI BgIII<br>~~~~~<br>~~~~~                                                                                                                                                             |
| 4761 | GATCTGGGG CTGGGTCTC CGGGCGTGAAG TCGGCCCGGA TCCTCGGGG GAATGGGGCT CTCGGATGTA<br>CTAGAGGGCC GACCGCAGAG GCCCGCACTC AGCCGGGCC AGGAGGGCC CTTACCCGA GAGCCTACAT<br>BglII BglIII<br>~~~~~<br>~~~~~ |
| 4831 | GATCTTCTTT CTTTCTCTT TTGTGGTAG AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTTCTT<br>CTAGAAGAAA GAAAGAAGAA AAACACCATC TAAACTTAG GGAGTGTAA CAAGTAGCCA TCAAAAAGAA                                    |
| 4901 | TTCATGATT GTGACAATG CAGGCCCTCGTG CGGAGCTTT TTGTAGGTAG<br>AAGTACTAAA CACTGTTAC GTGGAGCAC GCCTCGAAA AACATCCATC                                                                              |

FIG.\_43H



**FIG.\_44A**

ECORI

1 AATTCCACAA TGAACAATAA TAAGATTAAA ATAGCTTGCC CCCGTGTCAG CGATGGGTAT TTTTCTAGT  
TTAAGGTGTT ACTTGTATT ATTCTAATT TATCGAACGG GGGCAACGTC GCTACCCATA AAAAGATCA

71 AAAATAAAG ATAAACTTAG ACTCAAAACA TTTACAAAAA CAACCCCTAA AGTCCTAAAG CCAAAGTGC  
TTTTTATTTTC TATTGTGATC TGAGTTTGT AAATGTTTTT GTTGGGATT TCAAGGATTTC GGTTTCAAGC

141 TATGCAAGAT CCATAGCAAG CCCAGCCAA CCCAACCCAA CCCAACCCAC CCCAGTGCAG CCAAACGGCA  
ATACGTGCTA GGTATCGTTC GGTCGGGT GGTGGGGTT GGTGGGGGT GGTGACCGT GGTTGACCGT

211 AATAGTCTCC ACCCCCGGCA CTATCACCGT GAGTTGTCGG CACCAACCGCA CGTCTCGCAG CCAAACGGCA  
TTATCAGAGG TGGGGCCCGT GATACTGGCA CTCAACAGGC GTGGTGGCGT GCAGAGCGTC GGTTTGT

281 AAAAGAAAG AAAAGAAAG AAAAGAAAAA CAGCAGGTGG GTCCGGTGC TGGGGCCGG AAAAGGGAGG  
TTTTTCTTTC TTTTTTCTTCT TTTTCTTTT GTCGTCCACC CAGGCCAGC ACCCCCCGGCC TTTTCGCTCC

351 AGGATCGCGA GCAGCGACGA GGGCGGGCC TCCCTCGCT TCCAAGAAA CGCCCCCAT CGCCCACTATA  
TCCTAGCGCT CGTCGCTGGCT CGGGGGCGG AGGGAGGGCA AGGTTTCTTT GCGGGGGTA GCGGTGATAT

421 TACATACCCC CCCCTCTCCCT CCCATCCCCC CAACCCCTAAC ACCACCAACCA CCACCCACCTC CTCCCCCCC  
ATGTATGGGG GGGGAGAGGA GGGTAGGGG GTTGGGATGG TTGGTGGTGGT GGTGGTGGAG GAGGGGGAG

491 GCTGCCGGAC GACGAGGTCC TCCCCCTCC CGCTCCGGCG CCCTCCGGTAA CCACCCCGCC CCTCTCCCT  
CGACGGCCTG CTGCTCGAGG AGGGGGAGG GGGAGGGCC GGGGGCCATT GGTGGGGGG GGAGGGAGA

561 TTCTTCTCC GTTTTTTT TGTTCTCGGT CTGATCTTT GGCCTGGTA GTTGGTGG GCGAGAGGG  
AAGAAAGAGG CAAAGAAA AGCAGAGCCA GAGCTAGAAA CGGGAAACCAT CAAACCCACC CGCTCTCGCC

631 CTTCGTCGCC CAGATCGGTG CGGGGGAGGG GCGGGATCTC GCGGCTGGCG TCTCCGGCG TGAGTCGGCC  
GAAGCAGCGG GTCTAGCCAC GGGCCCTCCC CGCCCTAGAG CGCCGACCGC AGAGGCCCCG ACTCAGCGG

BamHI

701 CGGATCCTCG CGGGGAATGG GGCTCTCGGA TGTAGATCTT CTTTCTTTCT TCTTTCTTGTG GTAGAATTG  
GCCTAGGAGC GCCCCTTACCG CGAGAGGCCT ACATCTAGAA GAAAGAAAGA AGAAAACAC CATCTAAAC

FIG.-44B

|      |                                                                                                                                                               |         |  |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|---------|--|
| 771  | AATCCCTCAG CATTGTCAT CGGTAGTTT TCTTTTCATG ATTTGTGACA AATGCAGCCT CGTGGGGAGC TTAGGGAGTC GTAACAAGTA GCATCAAAA AGAAAAGTAC TAAACACTGT TTACGTGGA GCACGGCTCG         | HindIII |  |
| 841  | TTTTTGTAG GTAGAACGTT ACMATGGMCG TGGACACAGGA GGTSAACTTC GTSGCCTAACCG TCCTGATCGT AAAAACATC CATCTTCGAA TGKTACCKGC ACGTGTTCCCT CCASTGAAAG CASCGGATGG AGGACTAGCA   | NcoI    |  |
| 911  | SCTCGGCCTC CTCTTGCTCG TSTCCGCCAT GGAGCACGTT GACGCCAAGG CCTGCACCCCG CGAGTGGGGC SGAGCCGGAG GAGAACGGAG ASAGGGGGTA CCTCGTGCAC CTGCGTTCC GGACGTGGGM GCTCACGCG      | Not I   |  |
| 981  | AACCTCGGCT TCGGCATCTG CCCGGGGCC GCCTCCACGC AGGGCATTCTC CGAAGACCTC TACAGCCGTT TTGGAGCCGA AGGCCGTAGAC GGGCCGGCC CGGGGGTAGCTG TCCCCTGGAG GCTTCTGGAG ATGTCGGCAA   | SaI     |  |
| 1051 | TAGTCCAAT GGCCACTATC TCCCCAAGCTG CCTACGCCGA CCTGCGAAC ATTCCGTCGA CTATTATCAA ATCAGGTTTA CGGGTGTAG AGGGGTTGAC GGATGGGGCT GGACACGTTG TAAGGCAGCT GATAATAAGTT      | BamHI   |  |
| 1121 | GGGAGAGAAA ATTACAAATT CTCAAACTGA CATTAAACGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA CCCTCTCTTT TAAATGTAA GAGTTGACT GTATTGGCTT ACCTAGGAGG CGCTGCTGTC GTGTTTCTT       |         |  |
| 1191 | ATAATCACCG TCTTCCGGGG CACTGGTAGT GATAACGAAATC TACAACCTCGA TACTAACTAC ACCCTCACGC TATTAGTGGC AGAAGGCACC GTGACCATCA CTATGCTTAG ATGTTGAGCT ATGATTGATG TGGGAGTGC   |         |  |
| 1261 | CTTTCGACAC CCTTACCAAA TGCAACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGGT GAAAGGCTGAG GGATGGGTGTT ACGTGGTCCAA CACTTCATGT GCCACCTATA ATATAACCTA CCCAGAGGCA |         |  |

FIG.-44C

|      |             |             |             |             |             |             |             |
|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1331 | CCAGGACCAA  | GTCGAGTCGC  | TTGTCAAACA  | GCAGGGTAGC  | CAGTATCCGG  | ACTACGGCGT  | GACCGTGACC  |
|      | GGTCCTGGTT  | CAGCTCAGCG  | AACAGTTTGT  | CGTCAAATCG  | GTCAATGGCC  | TGATGGCGA   | CTGGCACTGG  |
| 1401 | GGCCACKCCC  | TGGGGCCTC   | CCTGGGGCA   | CTCACTGCCG  | CCCAGGCTGTC | TGCGACATAC  | GACAACATCC  |
|      | CCGGTGMGGG  | AGCCGGGAG   | GGACCGCCGT  | GAGTGAACGGC | GGGTGACAG   | ACGCTGTATG  | CTGTTGTTAGG |
| 1471 | GCCTGTACAC  | CTTCGGGAA   | CCCGCGAGCG  | GCAATCAGGC  | CTTCGGGTGCG | TACATGAACG  | ATGCCTTCCA  |
|      | CGGACATGTG  | GAAGCCGCTT  | GGCGCGTCGC  | CGTTAGTCCG  | GAAGGCCAGC  | ATGTACTTGC  | TACGGAAAGGT |
|      | XbaI        |             |             |             |             |             |             |
| 1541 | AGGCCTCGAGC | CCAGATACGA  | CGCAGTATT   | CCGGGTCACT  | CATGCCAACG  | ACGGCATCCC  | AAACCTGCC   |
|      | TCCGGAGCTCG | GGTCTATGCT  | GGGTCACTAAA | GGCCCAAGTGA | GTACGGTTGC  | TGCCGTAGGG  | TTGGACGGG   |
|      | NcoI        |             |             |             |             |             |             |
| 1611 | CCGGTGGAGC  | AGGGGTACGC  | CCATGGGGT   | GTAGAGTACT  | GGAGGGTGA   | TCCCTAACAGC | GCCCCAGAAC  |
|      | GGCCACCTCG  | TCCCCATGGCG | GGTACCCGCA  | CATCTCATGA  | CCTCGCAACT  | AGGAATGTGC  | CGGGTCTTGT  |
| 1681 | CATTGTCTG   | CACTGGGGAT  | GAAGTGCAGT  | GCTGTGAGGC  | CCAGGGGGAA  | CAGGGTGTGA  | ATAATGCCGA  |
|      | GTAAACAGAC  | GTGACCCCTA  | CTTCACGTCA  | CGACACTCCG  | GGTCCCGCCT  | GTCCCACACT  | TATTACGCGT  |
|      | NcoI        |             |             |             |             |             |             |
|      | SphI        |             |             |             |             |             |             |
| 1751 | CACGACTTAT  | TTTGGGATGA  | CGAGGGGGCGC | ATGCACCTGG  | CCGGTGGCGG  | CCGGGAAAC   | CACTGAAGGGA |
|      | GTGCTGAATA  | AAACCCCTACT | GCTCGCCCGG  | TACGGGACCC  | GGCCAGGGCC  | GGGCCCTTGT  | GTGACTTCCCT |
| 1821 | TGAGCTGTAA  | AGAAGCAGAT  | CGTTCAAACA  | TTGGCAATA   | AAAGTTCTTA  | AGATTGAATC  | CTGTTGCCGG  |
|      | ACTCGACATT  | TCTTCGTCTA  | GCAAGTTGT   | AAACCGTTAT  | TTCAAGAAAT  | TCTAACTTAG  | GACAACGGCC  |
| 1891 | TCTTGGGATG  | ATTATCATAT  | AATTCTGT    | GAATTACGTT  | AAGGATGTAA  | TAATTAACAT  | GTAATGCATG  |
|      | AGAACGCTAC  | TAATAGTATA  | TTAAAGACAA  | CTTAATGCAA  | TTCGTACATT  | ATTAATGTA   | CATTACGTAC  |
| 1961 | ACGTTATTAA  | TGAGATGGGT  | TTTTATGATT  | AGAGTCCCGC  | AATTATACAT  | TTAATACGCG  | ATAGAAAACA  |
|      | TGCAATAAT   | ACTCTACCCA  | AAAATACTAA  | TCTCAGGGCG  | TTAATATGTA  | AATTATGCC   | TATCTTTGTT  |

FIG.\_4D

09 09 01 2009 07 02 02

112 / 154

|      |                                                                                                                                                                     | XbaI   | XbaI   | XbaI   | XbaI   | XbaI   | XbaI   |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|--------|--------|--------|--------|--------|
|      |                                                                                                                                                                     | BssHII | BssHII | BssHII | BssHII | BssHII | BssHII |
| 2031 | AAATATAGCG CGCAAACTAG GATAAATTAT CGCGCGGGT GTCATCTATG TTACTAGATC GATAAGCTC<br>TTTATATCGC GCGTTTGATC CTATTAATA CGCGCGGCCA CAGTAGATAC AATGATCTAG CTATTCGAAG           |        |        |        |        |        |        |
|      |                                                                                                                                                                     | XbaI   | XbaI   | XbaI   | XbaI   | XbaI   | XbaI   |
| 2101 | TAGAGGGCC GGTGGAGCTC CAATTGGCCC TATAAGTGAATG CGTATTAAGT CGGCTCACTG GCGGTGCGTT<br>ATCTGGCGG CCACCTCGAG GTTAAGCGGG ATATCACTCA GCATAATGCG CGCAGGTGAC CGGCAGCAGAAA      |        |        |        |        |        |        |
| 2171 | TACAACGTG TGACTGGAA AACCCCTGGG TTACCCAACT TAATCGCCTT GCAGGCACATC CCCCTTTGCG<br>ATGTTGCAGC ACTGACCTT TTGGGACCCG AATGGTTGA ATTAGGGAA CGTCGTTAG GGGAAAGCG              |        |        |        |        |        |        |
| 2241 | CAGCTGGGT AATAGGGAAAG AGGCCGCAC CGATCGCCCT TCCCCAACAGT TGCGCAGGCCT GAATGGCGAA<br>GTCGACCGCA TTATCGCTTC TCAGGGCGTG GCTAGGGGA AGGGTTGTC ACCGCTACAC CTTACCGCTT         |        |        |        |        |        |        |
| 2311 | TGGGACGGC CCTGTAGGG CGCATTAAAGC GCGGGGGTG TGGTGGTTAC GCGCAGGGTG ACCGCTACAC<br>ACCCTGGCG GGACATGGC GGTAATTGCG CGGTAAATTGCG CGCCGCCAC ACCACCAATG CGCGTGCAC TGGCGATGTG |        |        |        |        |        |        |
| 2381 | TTGCCAGGC CCTAGGGCC GCTCCCTTTGCT CGTCTTTCCC TTCCCTTTCTC GCCACGTTCTG CGGGCTTTCC<br>AACGGTCGGC GGATCGGGG CGAGGAAAGC GAAAGAAGGG AAGGAAGAG CGGTGCAAGC GGCAGAAAGG        |        |        |        |        |        |        |
| 2451 | CCGTCAAGCT CTAATCGGG GGCCTCCCTT AGGGTTCCGA TTAGTTGCTT TACGGCACCT CGACCCCCAAA<br>GGCAGTTCGA GATTTAGCCC CGAGGGAAA TCCCAAGGCT AAATCACGAA ATGCCGTGGA GCTGGGGTT          |        |        |        |        |        |        |
| 2521 | AAACTTGATT AGGGTGTGG TTCACTGTTAGT GGGCCATCGC CCTGATAGAC GTTTTGACGT CCTTTTGACGT<br>TTTGAACCTAA TCCCACTACC AAGTGCATCA CCCGGTAGCG GGACTATCTG CCAAAAGCG GAAACTGCA       |        |        |        |        |        |        |
| 2591 | TGGAGTCCAC GTTCTTAAT AGTGGACTCT TGTCCAAAC TGGAAACAACA CTCAACCTA TCTCGGTCTA<br>ACCTCAGGTG CAAGAAATTA TCACCTGAGA ACAAGTTTG ACCTTGTGT GAGTTGGAT AGAGCCAGAT             |        |        |        |        |        |        |
| 2661 | TTCTTTGAT TTATAAGGGG TTGGCCGAT TCCGGCTAT TGGTTAAAAA ATGAGGTGAT TAAACAAA<br>AAGAAAACTA AATATTCCCT AAAACGGCTA AAGCCGGATA ACCAAATTCTT TACTCGACTA ATTGTTTT              |        |        |        |        |        |        |

FIG.-44E

2731 TTTAACCGCA ATTAAACAA AATATAACG CTTACAATT AGGTGGCACT TTTCGGGAA ATGTGGCGGG  
AAATTGGCT TAAAATTGTT TATAATTGC GAATGTTAAA TCCACCGTGA AAAGCCCTT TACACGGGCC

2801 AACCCCTATT TGTGTTATT TTCTAAATACA TTCAAAATATG TATCCGCTCA TGAGACAATA ACCCTGATAA  
TTGGGATAA ACAAAATAAA AGATTATGT AAGTTTATAC ATAGGGAGT ACTCTGTTA TGGGACTATT

2871 ATGCTCAAT AATATTGAAA AAGGAAGAGT ATGAGTATT AACATTTCCG TGTGCGCCCTT ATTCCCTTT  
TACGAAGTT TATAACTTT TTCCCTTCTCA TACTCATAAG TTGTAAGGC ACAGGGAA TAAGGGAAAA

2941 TTGGGCATT TTGCGCTTCCT GTTTTGGCTC ACCCAGAAC GCTGGTGAAG GTAAAGATG CTGAAGATCA  
AACGCCGTA AACGGAAAGGA CAAAACGAG TGGTCTTGC CGACCACTT CATTCTAC GACTTCTAGT

3011 GTTGGGTGCA CGAGTGGTT ACATCGAACT GGATCTCAAC AGGGGTAAGA TCCCTTGAGAG TTTTCGCCCC  
CAACCCACGT GCTCACCCAA TGAGCTTGA CCTAGAGTTG TGGTCTTGC AGGAACTCTC AAAAGGGGG

3081 GAAGAACGTT TTCCAATGAT GAGCACTTT AAAGTTCTGC TATGTGGCG GGTATTATGCC CGTATTGACG  
CTTCTTGCAA AAGGTTACTA CTCGTGAAAA TTCAAGACGG ATACACCGCG CCATAATTAGG GCATAACTGC

3151 CCGGGAAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAAGTCAC  
GGCCGGTTCT CGTTGAGCCA GGGCGTATG TGATAAAGAGT CTTACTGAAAC CAACTCATGA GTGGTCAGTG

3221 AGAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTAA TGCAGTGTGCTG CCATAACCCT GAGTGATAAC  
TCTTTCGTA GAATGCCCTAC CGTACTGTCA TTCTCTTAAT ACGTACGAC GGTATTGGTA CTCACTATG

3291 ACTGCGGCCA ACTTACTTCT GACAACGATC GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG  
TGACGCCGGT TGAATGAGA CTGTTGCTAG CCTCCTGGCT TCCTCGATTG GCGAAAAAAC GTGTTGTAC

3361 GGGATCATGT AACTCGCCTT GATCGTTGGG AACCGGAGCT GAATGAAAGCC ATACCAAACG ACGAGCGTGA  
CCCTAGTACA TTGAGCGGAA CTAGCAACCC TTGGCCTCGA CTTACTTCGG TATGGTTGC TGCTCGCACT

3431 CACCAAGATG CCTGTAGCAA TGGCAACAA GTTGCGCAA CTATTAACTG GCGAAACTACT TACTCTAGCT  
GTGGTGTAC GGACATCGTT ACGTTGTG CAAAGCGTT GATAATTGAC CGCTTGATGA ATGAGATCGA

3501 TCCCGGCAAC AATTAATAGA CTGGATGGAG GGGATAAAG TTGCAAGGACC ACTTCTGGC TCGGCCCTTC  
AGGGCGTGTG TAAATTATCT GACCTACCTC CGCCTATTTC AACGTCCTGG TGAAGACGG AGCCGGAAAG

|      |                          |                           |                           |                             |                           |                          |
|------|--------------------------|---------------------------|---------------------------|-----------------------------|---------------------------|--------------------------|
| 3571 | CGGCTGGCTG<br>GCCGACCGAC | GTATTATTGCT<br>CAAATAACGA | GATAAATCTG<br>CTATTAGAC   | GAGGCCGGTGA<br>CTCGGCCACT   | GGGTGGGTCT<br>CGGCCATAGT  | CGCGGTATCA<br>AACGTCGTGA |
| 3641 | GGGGCCAGAT<br>CCCCGGTCTA | GGTAAGCCCT<br>CCATTGGGA   | CCCGTATCGT<br>GGGCATAGCA  | AGTTATCTAC<br>TCAATTAGATG   | ACGACGGGAA<br>TGCTGCCCT   | GTCAGGCAAC<br>CAGTCGTTG  |
| 3711 | CGAAATAGAC<br>GCTTTATCTG | AGATCGCTGA<br>TCTAGGGACT  | GATAGGTGCC<br>CTATCCACGG  | TCACTGATTAA<br>AGTGAACATAAT | AGCATGGTA<br>TCGTAACCAT   | TATGGATGAA<br>ATACCTACTT |
| 3781 | CATATATACT<br>GTATATATGA | TTAGATTGAT<br>AATCTAACTA  | TTAAAACCTC<br>AATTGTTGAAG | ATTTTAAATT<br>TAAAAATTAA    | AAAAGGATC<br>ATTTCCTAG    | TAAGGAGTC<br>ATCCACTTCT  |
| 3851 | TAATCTCATG<br>ATTAGAGTAC | ACCAAAATCC<br>TGTTTATCG   | CTTAACGTGA<br>GAATTGCACT  | GTTTTCGTTC<br>CAAAGCAAG     | CACTGAGCGT<br>GTGACTCGCA  | CACTGAGCGT<br>GTCTGGGCA  |
| 3921 | AAAGGATCTT<br>TTTCCTAGAA | CTTGAGATCC<br>GAACCTCTAGG | TTTTTTCTG<br>AAAAAAAGAC   | CGCGTAATCT<br>GCGCATTAGA    | GCTGCTTGCA<br>CGACGAACGT  | CCACCGCTAC<br>TCTTTTCTAG |
| 3991 | CAGCGGTGGT<br>GTCGCCACCA | TGTTTTGCCG<br>ACAAACGGC   | GATCAAGAGC<br>CTAAGTCTCG  | TACCAACTCT<br>ATGGTTGAGA    | TTTTCCGAAG<br>AAAAGCTTC   | GTAACTGGCT<br>CATTGACCGA |
| 4061 | GCAGATACCA<br>CGTCTATGGT | AATACTGTC<br>TTATGACAGG   | TTCTAGTGT<br>AAGATCACAT   | GCCGTAGTTA<br>GGCATCAAT     | GGCCACCACT<br>CCGGGGTGA   | TCTAGGACCG<br>AGTTCTGAG  |
| 4131 | CCTACATACC<br>GGATGTATGG | TCGCTCTGCT<br>AGCGAGACGA  | AATCCCTGTT<br>TTAGGACAAAT | CCAGTGGCTG<br>GGTACCCGAC    | CGATAAGTGTG<br>GACGGTCACC | TGTCTTACCG<br>GCTATTAGGC |
| 4201 | GGTTGGACTC<br>CCAACCTGAG | AAGACGATAG<br>TTCTGCTATC  | TTACCGGATA<br>AATGGCCTAT  | GGGGCAGGG<br>TCCGGTGCAC     | CGATAAGTGTG<br>CAGCCGACT  | CGTGCACACA<br>TGCCCCCAA  |
| 4271 | GCCCAGCTTG<br>CGGGTGAAC  | GAGCGAACGA<br>CTCGCTTGCT  | GGAGACGGATA<br>GGATGGGCT  | AGGGCGAGCG<br>TGACTCTATG    | ACGGGGGTG<br>GATGTGCAC    | CGACGTGTGT<br>TCGATACTCT |
| 4341 | CTTCCCGAAG<br>GAAGGGCTTC | GGAGAAAGGC<br>CCTCTTCCG   | GGACAGGTTA<br>CCTGTCCATA  | CTACAGGGTGG<br>TGACTCTATG   | AGCTATGAGA<br>GATGTGCAC   | AAGGGGAG<br>TTCGCGGTG    |

FIG.-44G

4411 AGCTTCCAGG GGGAAACGCC TGGTATCTTT ATAGTCCCTGT CGGGTTTCGC CACCTCTGAC TTGAGCGTTC  
TCGAAGGTCC CCCTTTGGGG ACCATAGAAA TATCAGGACA GCCCCAAAGCG AACTCGCAGC

4481 ATTTTGTGA TGCTCGTAG GGGGGGGAG CCTATGGAA AACGCCAGCA ACGCCCTT TTACGGTT ACGGGGCTT TTTACGGTT  
TAAAACACT ACAGGCAGTC CCCCGCCCT GGATACCTTT TGCGGGTGT TGCGGGGAA AAATGCCAAG

4551 CTGGCCTTT GCTGGCCTTT TGCTCACATG TTCTTTCTGT CGTTATCCCC TGATTCTGTG GATAACCGTA  
GACCGAAAA CGACCGAAA ACAGGTGTAC AAGAAAGGAC GCAATAGGGG ACTAAAGACAC CTATTGGCAT

4621 TTACCGCCTT TGAGTGAAGCT GATAACCGCTC GCCGCAAGCCG AACGACCGAG CGCAGGGAGT CAGTGAAGGA  
AATGGGGAA ACTCACTCGA CTATGGGAG CGGCTGGCTC GCGTCGCTCA GTCACTCGCT

4691 GGAAGGGAA GAGGCCCAA TAGCCAAACC GCCTCTCCCC GCGCGTTGGC CGATTCAATA ATGCAGCTGG  
CCTTCGCCTT CTGGGGTT ATGGGTTGG CGGAGGGG CGCGCAACCG GCTAAAGTAAT TACGTCGACC

4761 CACGACAGGT TTCCCGACTG GAAAGGGGC AGTGAAGGCA ACGCAATAA TGTGAGTTAG CTCACTCATT  
GTGCTGTCCA AAGGGTGTAC CTTCGCCCG TCACTCGGT TGGGTAAATT ACACTCAATC GAGTGAAGTAAT

4831 AGGCACCCCA GGCTTTACAC TTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAAGG GATAACAATT  
TCCGGGGGT CGAAATGTG AAATACGAAG GCCGAGGATA CAACACACCT TAACACTCGC CTATTGGTAA

BssHII  
~~~~~

4901 TCACACAGGA AACAGCTATG ACCATGATA CGCCAAGGCG GCAATTAAACC CTCACTAAAG GAAACAAAG
AGTGTGTCCCT TTGTGATAAC TGGTACTAAT GCGGTTCGCG CGTTAAATTGG GAGTGATTTC CCTGTGTTTC

ECOR

4971 CTGG
GACC

FIG._44H

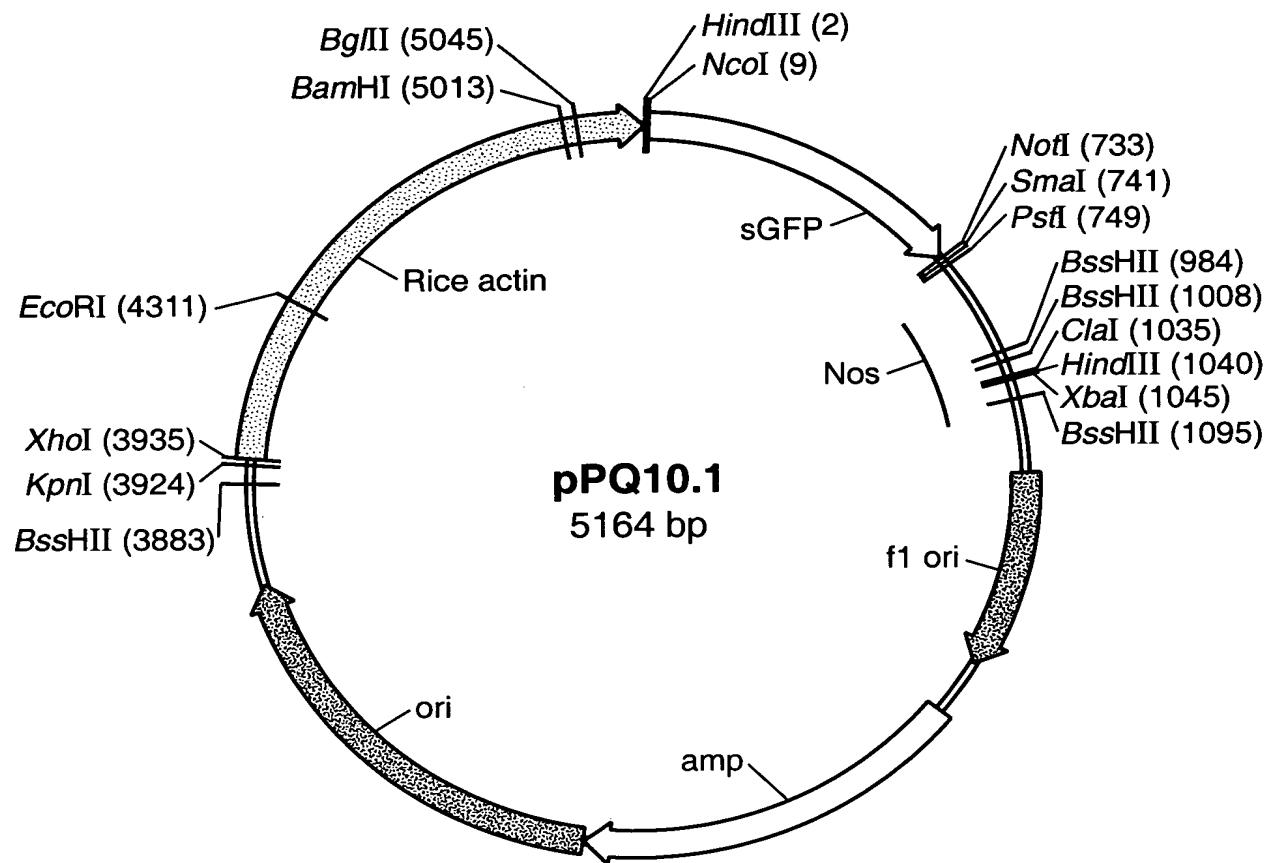


FIG.-45A

HindIII NcoI

1 AAGCTTACCA TGGTGAGCAA GGGGAGGAG CTGTTCACCG GGGTGGTGC CACCTGGCC GAGCTGGACG
TTCGAATGGT ACCACTCGTT CCCGGCTCCCTC GACAAGTGGC CCCACACGG GTAGGACCAAG CTCGACCTTGC
71 GCGACCGTGA CGGCCAACAG TTCAAGGTGT CCGGGAGGG CGAGGGGAT GCCACCTAAG GCAAGCTGAC
CGCTGCACTT GCGGGTTTC AAGTCGCACA GGCGGCTCCC GCTCCCGCTA CGGGGGATGC CGTTCGACTG
141 CCTGAAAGTTC ATCTGCACCA CGGGCAAGGT GCCCGTGCCT TGGCCACCCC TCGTGACCCAC CTTCACCTAC
GGACTTCAAG TAGACGTGGT GGCGGTTGA CGGGCACGGG ACCGGGTGG ACCACTGGTG GAAAGTGGATG
211 GGGGTGCAGT GCTTCAGCCG CTACCCGAC CACATGAAGC AGCACGACTT CTTCAAGTCC GCCATGCCCG
CCGCACGTCA CGAACGTGGC GATGGGGCTG GTGTACTTCG TGTTGCTGAA GAAGTTCAGG CGGTACGGGC
281 AAGGCTACGT CCAGGAGGGC ACCATCTCT TCAAGGACGA CGGCAACTAC AAGACCGGG CGGAGGTGAA
TTCCGATGCA GGTCCCTCGCG TGTTAGAAGA AGTTCTCTGCT GCCGTGTGATG TTCTGGGGC GGCTCCACTT
351 GTCGAGGGC GACACCTGG TGAAACCGAT CGAGCTGAAG GGCAATCGACT TCAAGGAGGA CGGCAACATC
CAAGCTCCCG CTGGGGACC ACTTGGGCTA GCTCGACTTC CCGTAGCTGA AGTTCCCTCCT GCCGTTGTAG
421 CTGGGGCACA AGCTGGAGTA CAACTACAAAC AGCCACAAACG TCTATATCAT GGCGGACAAG CAGAAGAACG
GACCCGTGT TCGACCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA CCGGCTGTTC GTCTTCTTGC
491 GCATCAAAGT GAACTTCAAG ATCCGCCACA ACATCGAGGA CGGCAGCGTG CAGCTGCCG ACCACTACCA
CGTAGTTCCA CTTGAAGTTTCTAGGGGTGT TGTAGCTCCT GCGGTGCCAC GTCGAGCGGC TGGTGATGTT
561 GCAGAACACC CCCATGGCG ACGGCCCCGT GCTGCTGCC GACAACCACT ACCTGAGGCAC CCAGTCCGCC
CGTCTGTGG GGGTAGCCGC TGCGGGGCA CGACGACGGG CTGTTGGTGA TGGACTCGTG GGTCAGGGG
631 CTGAGCAAAG ACCCAAACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT CGTGACCGCC GCGGGGATCA
GAATCGTTTC TGGGGTTGCT CTTCGCGCTA GTGTACCAAG ACGACACTCAA GCACGGGGC CGGCCTAGT

FIG._ 45B

118 / 154

| | | SmaI | NotI | PstI | | |
|------|-------------|-------------|------------|-------------|-------------|--------------|
| 701 | CTCACGGCAT | GGACGAGCTG | TACAAGTAAA | GCGGCCGCC | GGGCTGCAGG | AAGGATGAGC |
| | GAGTGGCGTA | CCTGCTCGAC | ATGTTCAATT | CGCCGGGGG | CCCGACGTCC | CTTGGGTGAC |
| 771 | TGTAAGAAG | CAGATCGTTTC | AAACATTGG | CAATAAAGTT | TCTTAAGATT | GAATCCTGTT |
| | ACATTCTTC | GTCTAGCAAG | TTGTAAACC | GTTATTCAA | AGAATCTAA | CTTAGGACAA |
| 841 | CGATGATTAT | CATAAAATT | CTGTTGAATT | ACGTTAAGCA | TGTAATAATT | AACATGTAAT |
| | GCTACTAATA | GTATATAAA | GACAACCTAA | TGCAATTCTGT | ACATTATAA | TTGTACATTA |
| 911 | ATTATGAGA | TGGTTTTTA | TGATTAGAGT | CCCGCAATTAA | TACATTTAAT | ACCGATAAGA |
| | TAATACTCT | ACCCAAAAAT | ACTAATCTCA | GGGGTTAAT | ATGTAATTA | TGGCTATCT |
| | | | | | XbaI | |
| | | | | | BssHII | |
| | | | | | | Clai HindIII |
| 981 | TAGCGCGCAA | ACTAGGATAA | ATTATCGGCC | GCGGTGTCAT | CTATGTTACT | AGATCGATAA |
| | ATCGCGCGTT | TGATCCTATT | TAATAGCGCG | CGCACACGTA | GATAAAATGA | TCTAGCTATT |
| | | | | | BssHII | |
| 1051 | CGGCGGTGG | AGCTCCAATT | CGCCCTATAG | TGAGTCGTAT | TACGCGCGCT | CACTGGCGT |
| | GGGGCCACC | TCGAGGTAA | GGGGATATC | ACTCAGGATA | ATGGCGCGA | CGAAAATGTT |
| 1121 | CGTCGTGACT | GGGAAACCC | TGGCGTTACC | CAACTTAATC | GCCTTGCAGC | ACATCCCCCT |
| | GCAGGCACTGA | CCCTTTGGG | ACGGCAATGG | GTTGAATTAG | CGGAACGTCTG | TGTAGGGGA |
| 1191 | GGCGTAATAG | CGAAGAGGCC | CGCACCGATC | GCCCTTCCA | ACAGTTGCAG | AGCCTGAAATG |
| | CCGCATTATC | GCTTCTCGG | GGTGGCTAG | CGGGAAAGGGT | TGTCAAACGCG | TCGGAACTTAC |
| | | | | | | CGCTTACCT |

FIG._45C

1261 CGGCCCTGT AGGGGGCAT TAAGGGGCC GGGTGTGGT GTTACGGCA GCGTGAACCGC TACACTTGGCC
GGCGGGACA TCGCGGCCGTAA TCGCGGCCGCCC CCCACACCAC CAATGCCGT CGCACTGGCG ATGTGAACGG

1331 AGGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTGCCAC GTTCCGGGC TTTCGGCG
TCGGGGATC GCGGGGAGG AAAGCGAAAG AAGGGAAAGG AAGAGGGTCA AAGAGGGCG AAAGGGGCG

1401 AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT TCCGATTAG TGCTTACGG CACCTCGACCC CCAAAAGCT
TTCGAGATT AGCCCCCGAG GAAAATCCCA AGGCTTAATC ACGAAATGCC GTGGAGCTGG GTTGGTTGAG

1471 TGATTAGGGT GATGGTTCAC GTAGGGGCC ATCGCCCTGA TAGACGGTT TTGCCCCCTT GACGTTGGAG
ACTAATCCCA CTACCAAGTG CATCACCCGG TAGCGGGACT ATCTGCCAAA AAGGGGAAA CTGCAACCTC

1541 TCCACGGTCT TTAATATGGG ACTCTTGTTC CAAACACTGGAA CAACACTCAA CCCTATCTCG GTCTTATTCTT
AGGTGCAAGA ATTATCACC TGAGAACAG GTTGTACCTT GTTGTGAGTT GGGATAGAGC CAGATAAGGA

1611 TTGATTATA AGGGATTTG CGGATTTGGG CCTATTGGTT AAAAATGAG CTGATTAAAC AAAAATTAA
AACTAATAT TCCCTAAAC GGCTAAAGCC GGATAACCAA TTTTTACTC GACTAAATTG TTTTTAAATT

1681 CGCGAATT TACAAATAT TAACGCTTAC AATTAGGTG GCACTTTTCG GGGAAATGTT CGGGGAACCC
GGCGTTAAA TTGTTTATA ATGCGGAATG TAAATCCAC CGTGAAGAC CCCTTTACAC GCGCCTTGGG

1751 CTATTGTT ATTTCCTAA ATACATTCAA ATATGTATCC GCTCATGAGA CAATAACCCCT GATAAAATGCT
GATAAACAA TAAAAGATT TATGTAAGTT TATACATAGG CGAGTACTCT GTTATTGGGA CTATTTACGA

1821 TCAATAATAT TGAAAAGGA AGAGTATGAG TATTCAACAT TTCCGTGTCG CCCTTATTCC TTGTTTTCGCG
AGTTATTATA ACTTTTCCCT TCTCATACTC ATAAGTTGTA AAGGCACAGC GGGAAATAAGG GAAAAAACGC

1891 GCATTGTC TTCCGTGTT TGCTCACCA GAAACGCTGG TGAAAGTAAAG AGATGCTGAA GATCAGTTGG
CGTAAACCGG AAGGACAAAA AGGAGTGGGT CTTGCGACCC ACTTCATT TCTACGACTT CTAGTCACCC

1961 GTGCACGGGT GGGTTACATC GAACTGGATC TCACACAGGG TAAGATCCTT GAGAGTTTC GCCCCGAAAGA
CACGTGCTCA CCCAATGTTAG CTTGACCTAG AGTGTGCGCC ATTCTAGGAA CTCTCAAAAG CGGGGCTTCT

2031 ACGTTTCCA ATGATGAGCA CTTTAAGT TCTGCTATGT GGGGGGGTAT TATCCCGTAT TGACGGGGGG
TGCAAAAGGT TACTACTCGT GAAAATTCA AGACGATACA CGCGCCATA ATAGGGCATA ACTGCGGGCCC

| | | | | | | | |
|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 2101 | CAAGAGAAC | TCGGTGGCCG | CATACACTAT | TCTCAGAAATG | ACTTGGTTGA | GTACTCACCA | GTCACAGAAA |
| | GTTCCTCGTTG | AGCCAGGGC | GTATGTGATA | AGAGTCTTAC | TGAACCAACT | CATGAGTGGT | CAGTGTCTT |
| 2171 | AGCATCTTAC | GGATGGCATG | ACAGTAAGAG | AATTATGCAAG | TGCTGCCATA | ACCATGAGTG | ATAACACTG |
| | TCGTAGATG | CCTAACCGTAC | TGTCATTCTC | TTAATACGTC | ACGACGGTTAT | TGGTACTCAC | TATTGTGACG |
| 2241 | GGCCAACTTA | CTTCTGACAA | CGATGGAGG | ACCGAAGGGAG | CTAACCGGCTT | TTTGCACAA | CATGGGGGAT |
| | CCGGTTGAAT | GAAGACTGTT | GCTAGCCCTC | TGGCTTCCTC | GATTGGCAA | AAAAGTGTGTT | GTACCCCTA |
| 2311 | CATGTAACTC | GCCTTGATCG | TGGGAACCG | GAGCTGAATG | AAGCCATACC | AAACGACCGAG | CGTGACACCCA |
| | GTACATTGAG | CGGAACTAGC | AACCCCTGGC | CTCGACTTAC | TTCGGTATGG | TTGGCTGCTC | GCACGTGTGGT |
| 2381 | CGATGCCCTGT | AGCAAATGGCA | ACAAACGTTGC | GCAAACATT | AACTGGCGAA | CTACTTACTC | TAGCTTCCCCG |
| | GCTACGGACA | TCGTTACCGT | TGGTCAACG | CGTTTGATAA | TTGACCGCTT | GATGAATGAG | ATCGAAGGGC |
| 2451 | GCAACAATT | ATAGACTGGA | TGGAGGGGA | TAAAGTTGCA | GGACCACTTC | TGGGGCTCGGC | CCTTCCGGCT |
| | CGTTGTTAAT | TATCTGACCT | ACCTCCGGCT | ATTCAACGT | CCTGGTGAAG | ACGGGAGCCG | GGAAAGGGCGA |
| 2521 | GGCTGGTTA | TTGCTGATAA | ATCTGGGCC | GGTGAGCGTG | GGTCTCGGGG | TATCATTGCA | GCACGTGGGGC |
| | AACGACATT | TAGACCTCGG | CCACTCGCAC | CCAGAGGGCC | CCAGAGGGCC | ATAGTAACGT | CGTGACCCCG |
| 2591 | CAGATGGTAA | GCCCTCCCGT | ATCGTAGTTA | TCTACACGAC | GGGGAGTCAG | GCAACTATGG | ATGAACCGAAA |
| | GTCTTACCAT | CGGGAGGGCA | TAGCATCAAT | AGATGTGCTG | CCCCTCAGTC | CGTTGATACC | TACTTGCTTT |
| 2661 | TAGACAGATC | GCTGAGGATAG | GTGCCTCACT | GATTAAGCAT | TGGTAACTGT | CAGACCAAGT | TTACTCATAT |
| | ATCTGTCTAG | CGACTCTATC | CACGGAGTGA | CTAATTGATA | ACCATTGACA | GTCTGGTCA | AATGAGATA |
| 2731 | ATACCTTAA | TGATTAAA | ACTTCATTT | TAATTAAA | GGATCTAGGT | GAAGATCCTT | TTTGATAATC |
| | TATGAAATCT | AACTAAATT | TGAAGTAAA | ATAAATTT | CCTAGATCCA | CTTCTAGGAA | AAACTATTAG |
| 2801 | TCATGACCAA | AATCCCTTAA | CGTGAGTTT | CGTTCCACTG | AGGGTCAGAC | CCCGTAGAAA | AGATCAAAGG |
| | AGTACTGGTT | TTAGGGAAATT | GCACCTCAA | GCAGGGTGAAC | TCGGCAGTCTG | GGGCATCTTT | TCTAGTTCC |
| 2871 | ATCTCTTGA | GATCCCTTTT | TTCTGGGGT | AATCTGCTGC | TTGCAACAA | AAAACCACC | GCTACCAAGGG |
| | TAGAAGAACT | CTAGGAAAAA | AAGACGGCA | TTAGACGACG | AACGTTGGTGG | CGATGGTCGC | |

| | | | | | | | |
|------|-------------|-------------|-------------|-------------|--------------|--------------|-------------|
| 2941 | GTGGTTGTT | TGCCGGATCA | AGAGCTACCA | ACTCTTTTC | CGAAAGGTAAC | TGGCTTCAGC | AGAGCGCAGA |
| | CACCAAACAA | ACGGCCCTAGT | TCTCGATGGT | TGAGAAAAAG | GCTTCCATTG | ACCGAAGTCG | TCTCGGGTCT |
| 3011 | TACCAAATACT | TGTCCCTCTA | GTGTTAGCCGT | AGTTAGGCCA | CCACTTCAAG | AACTCTGTAG | CACCGCCCTAC |
| | ATGGTTTATG | ACAGGAAGAT | CACATCGGCA | TCAATCCGGT | GGTGAAGTTC | TGAGGACATC | GTGGCGGATG |
| 3081 | ATACCTCGCT | CTGCTTAATCC | TGTTACCACT | GGCTGCTGCC | AGTGGCGATA | AGTCGTTGCTCT | TACCGGGTTG |
| | TATGGAGGGA | GACGATTAGG | ACAATGGTCA | CCGACGACGG | TCACCGCTAT | TCAGCACAGA | ATGGCCCAAC |
| 3151 | GAATCAAGAC | GATAGTTAC | GGATAAAGGG | CAGCGGTCCG | GCTGAACGGG | GGGTTCTGTC | ACACAGCCCCA |
| | CTGAGTTCTG | CTATCAATGG | CCTATTCCGC | GTGCCCCAGCC | CGACTTGGCC | CCAAAGCACG | TGTGTGGGT |
| 3221 | GCTTGGAGCG | AACGACCTAC | ACCGAACTGCA | GATACTTACA | GCGTGAAGCTA | TGAGAAAGCG | CCACGCTTCC |
| | CGAACCTCGC | TTGCTGGATG | TGGCTTGACT | CTATGGATGT | CGCAACTCGAT | ACTCTTCGC | GGTGCAGAAGG |
| 3291 | CGAAGGGAGA | AAGGGGGACA | GGTATCCGGT | AAGGGCAGG | GTCCGAACAG | GAGAGGCAC | GAGGGAGCTT |
| | GCTTCCCTCT | TTCCGGCTGT | CTATAGGCCA | TTCGGCTGCC | CAGCCTTGTGTC | CTCTCGCGTG | CTCCCTCTGAA |
| 3361 | CCAGGGGGAA | ACGCCTGGTA | TCTTTATAGT | CCTGTGGGT | TTCCGCCACCT | CTGACTTGAG | CGTCGATTCTT |
| | GGTCCCCCCTT | TGGGGACCAT | AGAAATATCA | GGACAGGCCA | AAGGGTGGAA | GACTGAACTC | GCAGCTAAAA |
| 3431 | TGTGATGCTC | GTCAGGGGG | CGGAGCCAT | GGAAAACGC | CAGCAACCGC | GCCTTTTAC | GGTTCCCTGGC |
| | ACACTACGAG | CAGTCCCCC | GCCTCGGGATA | CTTTTGGC | GTCCGGCGC | CGGAAAATG | CCAAGGACCG |
| 3501 | CTTTGCTGG | CCTTTGCTC | ACATGTTCTT | TCTCTGGTAA | TCCCCCTGATT | CTGTGGATAAA | CGTATATAACC |
| | GAAAACGACC | GGAAAACGAG | TGTACAAGAA | AGGACGCAAT | AGGGGACTAA | GACACCTATT | GGCATAATGG |
| 3571 | GCCTTTGAGT | GAGCTGATACT | CGCTCGCCGC | AGCCGAACGA | CCGAGGCCAG | CGAGTCAGTG | AGCGAGGAAG |
| | CGAAAAACTCA | CTCGACTATG | GGAGGGGGC | TGGCTTGTCT | GGCTCGCGTC | GCTCAGTCAC | TCGCTCTTC |
| 3641 | CGGAAGAGCG | CCCAATACGC | AAACCGCCTC | TCCCCGGCG | TTGGCCGATT | CATTAATGCA | GCTGGCACGA |
| | GCCTTCTCGC | GGGTTATGCG | T'TTGGGGAG | AGGGGGCGC | AACCGGCTAA | GTAATTACGT | CGACCGTGCT |
| 3711 | CAGGTTTCCC | GACTGAAAG | CGGGCAGTGA | GCGCAACGCA | ATTAATGTA | GTAGCTCAC | TCATTAGGCA |
| | GTCCAAAGGG | CTGACCTTTC | GCCCCGTCACT | GGCGTTGCGT | TAATTACACT | CAATCGAGTG | AGTAATCCGT |

| | | KpnI | KpnI | |
|------|--|--------|--------|-------|
| 3781 | CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGGTT GTGGGATTGT GAGCGGATAA CAATTCAACA GGGGTCCGAA ATGTGAAATA CGAAGGCCGA GCATACAACA CACCTAACCA CTCGCCTATT GTTAAAGTGT | BssHII | BssHII | |
| 3851 | CAGGAAACAG CTATGACCAT GATTACGCCA AGCGGCCAAT TAACCCCTCAC TAAAGGAAAC AAAAGCTGGG GTCCTTTGTC GATACTGGT CTAATGGGT TCGCGCGTTA ATTGGGAGTG ATTTCCTTG TTTTCGACCCC | XbaI | XbaI | |
| 3921 | TACCGGGCCC CCCCTCGAGG TCATTCCATAT GCTTGAGAAG AGAGTCGGGA TAGTCACAAA TAAACAAAG ATGGCCCGGG GGGGAGCTCC AGTAAGTATA CGAACTCTTC TCTCAGCCCT ATCAGGTTTT ATTTCGTTTC | | | |
| 3991 | GTAAGATTAC CTGGTCAAA GTGAAAACAT CAGTTAAAG GTGGTATAAG TAAATATCG TAAATAAAG CATTCAATG GACCAGTTT CACTTTGTA GTCAATTTC CACCATATTG ATTTTATAGC CATTATTTC | | | |
| 4061 | GTGGCCAAA GTGAAAATTAA CTCTTTCTA CTATTATAAA ATTGAGGAT GTTGGTCTGG TACTTTGATA CACCGGGTTT CACTTTAAT GAGAAAAGAT GATAATATTG TTAACTCCTA CAAACAGCC ATGAAAACAT | | | |
| 4131 | CGTCATTTTT GTATGAAATTG GTTTTTAAGT TTATTCCGCGA TTGGGAATG CATACTCTGTA TTGAGTCGG GCAGTAAAAA CATACTAAC CAAAAATTCA ATAAGGCCT AAACCTTTAC GTATAGACAT AAACTCAGCC | | | |
| 4201 | TTTTAAGTT CGTTGCTTTT GTAAATACAG AGGGATTGT ATAAGAAATA TCTTTAAAA ACCCATATGC AAAATTCAA GCAACGAAAA CATTATGTC TCCTAACAA TATTCTTAT AGAAATTTT TGGGTATACG | | | ECORI |
| 4271 | TAATTGACA TAATTGTA GAAAATATA TATTCAAGCG AATTCCACAA TGAACAAATAA TAAGATAAA ATTAACATGT ATTAACAACT CTTTTATAT ATAAGTCCGC TTAAGGTGTT ACTTGTATT ATTCTAATT | | | |
| 4341 | ATAGCTTGCC CCCGTTGCAG CGATGGGTAT TTTCCTAGT AAATAAAAG ATAAACTTAG ACTCAAAACA TATCGAACGG GGGCAACGTC GCTACCCATA AAAAGATCA TTTTATTTTC TATTGAAATC TGAGTTTGT | | | |
| 4411 | TTTACAAAA CAACCCCTAA AGTCCTAAAG CCCAAAGTGC TATGCCACGAT CCATAGCAAG CCCAGCCAA AAATGTTTTT GTGGGGATT TCAGGATTTC GGTTTCAAG ATACGTGCTA GGTATCGTT GGGTCGGTT | | | |

FIG.-45G

| | |
|------|---|
| 4481 | CCCAACCAA CCCAACCCAC CCCAGTGGAG CCAACTGGCA AATAGTCTCC ACCCCGGCA CTATCACCGT GGGTTGGTT GGGTTGGGT GGGTCAGTC GGTGTCAGTC GGTGACCGT TTATCAGAGG TGGGGCCGT GATA GTGGCA |
| 4551 | GAGTGTCCG CACCACCGCA CGTCTCGCAG CCAAAAAAAA AAAAGAAAG AAAAGAAA AAAAGAAAAA CTCAACAGGC GTGGTGGCGT GCAGAGCGTC GGTTTCTTTC TTTCAGCTTCTT TTTCAGCTTCTT |
| 4621 | CAGCAGGTGG GTCCGGGTG TGGGGGCGG AAAAGCGAGG AGGATCGCGA GCAGCGACGA GGCCCGGCC GTCGTCCACC CAGGCCAGC ACCCCGGCC TTTCAGCTCC TCCTAGCGCT CGTCGCTGCT CGGGCCGGG |
| 4691 | TCCCTCCGCT TCCAAAGAAA CGCCCCCAT CGCCCACTATA TACATAACCCC TACATAACCCC AGGGAGGCGA AGGTTTCTTT GCGGGGGGTA GCGGTGATAT ATGTTATGGGG GGGAGAGGA GGGTAGGGGG |
| 4761 | CAACCTTACCA ACCACCCACCA CCACCCACTC CTCCCCCCTC GCTGCCGGAC GACGGCTCC TCCCCCTCC GTTGGATGG TTGTTGGTGGT GGTGGTGGAG GAGGGGGAG CGACGGCCTG CTGCTGAGG AGGGGGAGG |
| 4831 | CCCTCCGGCG CCGCCGGTAA CCACCCGCC CCTCTCCTCT TTCTTCTCC GTGCTCTGGT GGGAGGGCCG GGGGCCATT GGTGGGGGG GGAGAGGAGA AAGAAAGAGG CAAAAAAA AGCAGAGCCA |
| 4901 | CTCGATCTT GGCTTGGTA GTTGGGGGG GCGAGAGCGG CTTCGTGCGCC CAGATCGGTG CGCGGGAGGG GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC GAAGCAGCGG GTCTAGCCAC GCGCCCTCCCC |
| | BamHI ~~~~~ |
| 4971 | GCGGGATCTC GCGGCTGGCG TCTCCGGCG TGAGTCGGCC CGGATCCTCG CGGGAAATGG GGCTCTCGGA CGCCCTAGAG CGCCGACCGC AGGGGCCGC ACTCAGCCGG GCCTAGGAGC GCCCCCTTACCG CGAGAGCCCT |
| | BglIII ~~~~~ |
| 5041 | TGTAGATCTT CTTTCTTCTTCT TCTTTGGTG GTAGAATTG AATCCCTCAG CATTGGTCAT CGGTAGTTT ACATCTAGAA GAAAGAAAGA AGAAAACAC CATCTAAAC TTAGGGAGTC GTAACAGTA GCCATCAAAA |
| 5111 | TCTTTCTATG ATTGTTGACA ATGCAAGCCT CGTGGGGAGC TTTTTGGTAG GTAG AGAAAAGTAC TAAACACTGT TTACGTGGA GCACGGCTCG AAAAACATC CATC |

FIG._ 45H

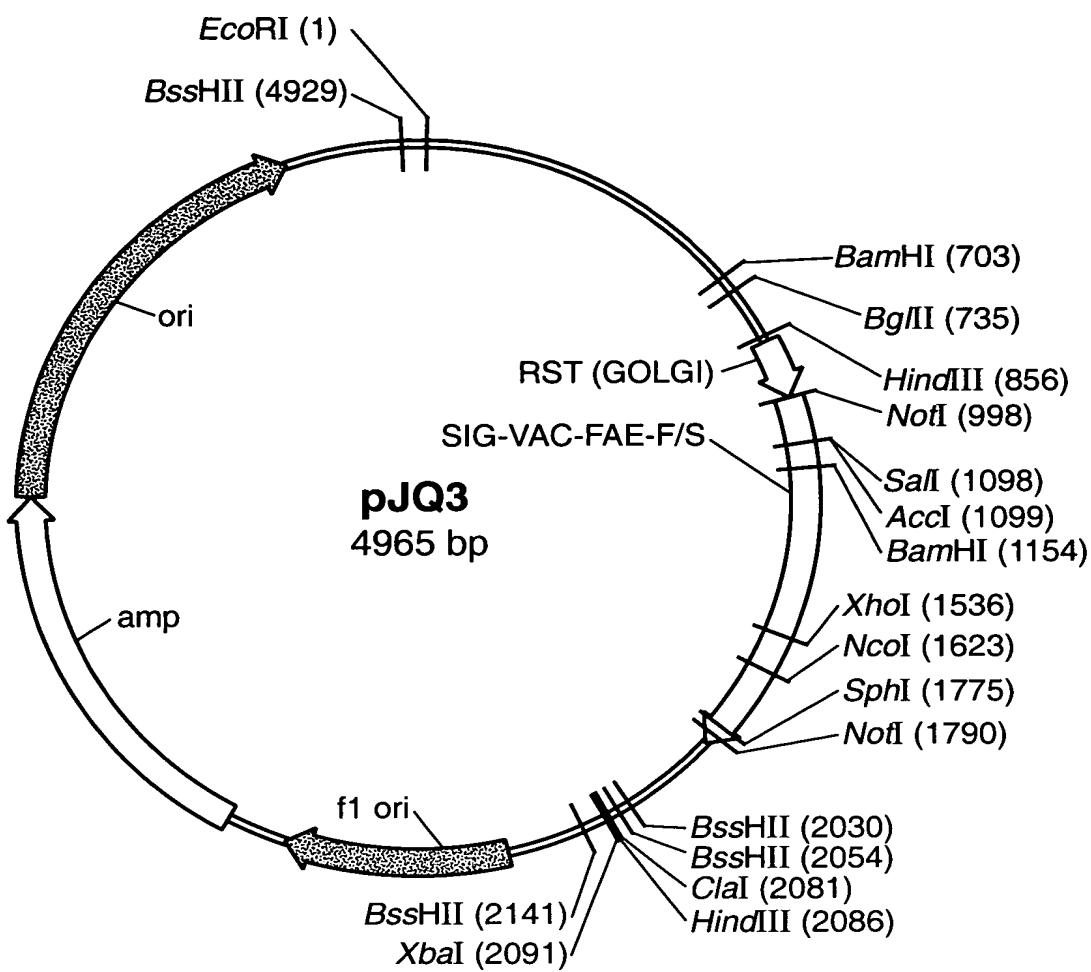


FIG._46A

ECORI

1 AATTCCACAA TGAACAATAA TAAGATTAAA ATAGCTTGGCC CCCGTGGCAG CGATGGTAT TTTTCTAGT
TTAAGGTGTT ACTTGTATT ATTCTAATT TATCGAACGG GGGAACGTC GCTACCCATA AAAAGATICA

71 AAAATAAAG ATAAACTTAG ACTCAAAACA TTTACAAAAA CAACCCCTAA AGTCCTAAAG CCCAAAGTGC
TTTATTTTC TATTGATC TGAGTTTGT AAATGTTTT GTGGGGATT TCAGGATTTC GGTTTTCACG

141 TATGCACGAT CCATAGCAAG CCCAGCCAA CCCAACCCAA CCCAACCCAC CCCAGTGCAG CCAACTGGCA
ATACGTGCTA GGTATGTTG GGTGGGGTT GGTTGGGT GGGTCACTG GTGTGACCGT

211 AATAGTCTCC ACCCCGGCA CTATCACCGT GAGTTGTCGG CACCACCGCA CGTCTGGCAG CCAAAAAAAA
TTATCAGAGG TGGGGCCGT GATACTGGCA CTCAACAGGC GTGGTGGCGT GCAGAGCGTC GTTTTTTTT

281 AAAAGAAAG AAAAGAAAG AAAAGAAAAA CAGCAGGTGG GTCCGGTGG TGGGGCGG AAAAGCAGGG
TTTTCTTTC TTTTTTTCT TTTTCTTTT GTCTCCACC CAGGCCAGC ACCCCCCGCC TTTTCGCTCC

351 AGGATGCGA GCAGCGAGA GGGCGGGCC TCCCTCCGCT TCCAAGAAA CGCCCCCAT CGCCCACTATA
TCCTAGCGCT CGTCGCTGGT CCGGGCCGG AGGGAGGCGA AGGTTCCTTT GCGGGGGTA GCGGTGATAT

421 TACATACCCC CCCCTCTCC CCTCATCCCC CAACCCCTACCC ACCACCAACCA CCACCCACCTC CTCCCCCTC
ATGTATGGGG GGGAGAGGA GGTAGGGGG GTGGGATGG TGGTGGTGGT GGTGGTGGAG GAGGGGGAG

491 GCTGCCGGAC GACGAGCTCC TCCCCCTCC CCCTCCGCC CGGCCGGTAA CCACCCGCC CCTCTCCTCT
CGACGGCCTG CTGCTCGAGG AGGGGGAGG GGGAGGGCC GGGGCCATT GTGGGGGG GAGAGGGAGA

561 TTCTTCTCC GTTTTTTTT TCGTCTCGGT CTCGATCTT GGCCTGGTA GTTTGGTGG GCGAGAGCGG
AAGAAAGGG CAAGAAAAA AGCAGAGCA GAGTAGAAA CGGGAACCAT CAAACCCACC CGCTCTCGCC

631 CTTCGTGGCC CAGATCGGTG CGGGGGAGG GCGGGATCTC GGGCTGGCG TCTCCGGCG TAGTGGGCC
GAAGCAGGG GTCTAGCCAC GGGCCCTCCC CGGCCCTAGAG CGGCCACCGC AGAGGGCCGC ACTCAGCCGG

BamHI

701 CGGATCCTCG CGGGGAATGG GGCTCTCGGA TGTAGATCTT CTTCCTTCT TCTTTTGTG GTAGAATTG
GCCTAGGAGC GCCCTTACCG CCGAGGCGT ACATCTAGAA GAAAGAAAGA AGAAAACAC CATCTAAAC

FIG._46B

771 AATCCCTCAG CATTGTTCAT CGGTAGTTT TCTTTTCATG ATTGTGACA AATGCAGCCT CGTGGGGAGC
TTAGGGAGTC GTAACAAGTA GCCATCAAAA AGAAAAGTAC TAAACACTGT TTACGTGGA GCACGGCTCG

HindIII

841 TTTTTTAG GTAGAACGTT ACCATGATCC ACACCAACCT CAAAGAAG TTCTCCCTCT TCATCCTCGT
AAAAAACATC CATCTTCGAA TGTTACTAGG TGTGGTTGGA GTTTTCTTC AAGAGGGAGA AGTAGGGAGCA

911 CTTCCCTCCTC TTCGCCGTGA TCTGGGTGTG GAAGAAGGGC TCCGACTACG AGGCCTCAC CCTCCAAGCC
GAAGGAGGAG AAGGGGCACT AGACGCACAC CTTCTTCCCCG AGGCTGATGC TCCGGGAAGTG GGAGGGTTCGG

Not I

981 AAGGAGTTCC AAATGGGGC CGCCTCCACG CAGGGCATCT CGAAGACCT CTACAGCCGT TTAGTCGAAA
TTCCCTCAAGG TTACCGCCG GCGGAGGTGC GTCCCGTAGA GGCTTCTGGA GATGTCGGCA AATCAGCTT

SaI

~~~~~

1051 TGGCCACTAT CTCCCAAGCT GCCTACGCCG ACCTGTGCAA CATTCCGTG ACTATTATCA AGGGAGAGAA  
ACGGGTGATA GAGGGTTCGA CGGATGCCG TGGACACGTT GTAAGGCAGC TGATAATAGT TCCCTCTCTT

BamHI

1121 AATTTACAAT TCTCAAACCTG ACATTAACGG ATGGATCCCTC CGCGACGACA GCAGCAAAGA ATAAATCACCC  
TTAATGTTA AGAGTTGAC TGATAATTGCC TACCTAGGAG GCGCTGCTGT CGTCGTTCT TTATTAATGTTG

1191 GTCTTCGGTG GCACTGGTAG TGATACGAA CTACAACACTCG ATACTAACTA CACCCCTCAG CCTTTCGACA  
CAGAAGGCAC CGTGACCATC ACTATGCTTA GATGTTGAGC TATGATTGAT GTGGGAGTGC GAAAGCTGT

1261 CCCTACCA AATGCAACGGT TTGTGAAGTAC ACGGTGGATA TTATATTGGA TGGGTCTCCG TCCAGGACCA  
GGGATGGTGT TACGTTGCCA ACAACTTCATG TGCCACCTAT AATATAACCT ACCCAGAGGC AGGTCCCTGGT

FIG.-46C

|      |                          |                           |                           |                          |                            |                           |                           |
|------|--------------------------|---------------------------|---------------------------|--------------------------|----------------------------|---------------------------|---------------------------|
| 1331 | AGTCGAGTCG<br>TCAGTCAGC  | CTTGTCAAAAC<br>GAACAGTTG  | AGCAGGTTAG<br>TGTCCAATC   | CCAGTATCCG<br>GGTCATAGGC | GAATCGGC<br>CTGATGGCG      | TGACCGTGAC<br>ACTGGCACTG  | CGGCCACKCC<br>GCCGGTGMGG  |
| 1401 | CTCGGGCCT<br>GAGCCGGGA   | CCCTGGGGC<br>GGGACCGCCG   | ACTCACTGCC<br>TGAGTGACGG  | GCCCAGCTGT<br>CGGGTCGACA | CTGGCACATA<br>GACGCTGTAT   | CGACAACATC<br>GCTGTTGTAG  | CGCCTGTACA<br>GGGGACATGT  |
|      |                          |                           |                           |                          |                            | XbaI                      |                           |
| 1471 | CCTTCGGCGA<br>GGAAGCGCT  | ACCGCGCAGC<br>TGGCGGTGCG  | GGCAATCAGG<br>CCGTTAGTTC  | CCTTCGGGTG<br>GGAAAGCGAG | GTACATGAAC<br>CATGTACTTG   | GATGCCCTTC<br>CTACGGAAAG  | AAGCCTCGAG<br>TTCGGAGGCTC |
| 1541 | CCCAGATAAG<br>GGGTCTATGC | ACGGCAGTATT<br>TGCGTCATAA | TCGGGGTCAC<br>AGGCCCAAGTG | TCATGCCAAC<br>AGTACGGTTG | GACGGCATCC<br>CTGCCGTAGG   | CAAACCTGCC<br>GTTTGGACGG  | CCGGTGGAG<br>GGGCCACCTC   |
|      |                          |                           |                           |                          |                            | NcoI                      |                           |
| 1611 | CAGGGGTACG<br>GTCCCCATGC | CCCATGGGG<br>GGGTACCGCC   | TGTTAGAGTAC<br>ACATCTCATG | TGGAGCGTTG<br>ACCTCGCAAC | ATCCTTACAG<br>TAGGAATGTC   | CGCCCAAGAAC<br>GGGGTCTTG  | ACATTGTCT<br>GTAAACAGA    |
| 1681 | GCACGGGGGA<br>CGTGACCCCT | TGAAGTGCAG<br>ACTTCACGTC  | TGCTGTGAGG<br>ACGACACTC   | CCCAGGGGG<br>GGGTCCCGCC  | ACAGGGGTG<br>TGTCACAC      | AATAATGCC<br>TTTACCGC     | ACACGACTTA<br>TGTGCTGAAT  |
|      |                          |                           |                           |                          |                            | SphI                      |                           |
| 1751 | TTTTGGGATG<br>AAAACCTAC  | ACGAGGGCG<br>TGCTCGCCG    | CATGGCACCTG<br>GTACGTGGAC | GCCGGTGC<br>CGGCCAGCGC   | GCCGGGAA<br>CGGGCCTTT      | CCACTGAAGG<br>GGTGACTTCC  | ATGAGGCTGTA<br>TACTCGACAT |
| 1821 | AAGAACGAGA<br>TTCTTCGTCT | TCTGTCAAAAC<br>AGCAAGTTG  | ATTGGCAAT<br>TAAACCGTTA   | AAAGTTCTT<br>TTCTAAAGAA  | AAGATGAAAT<br>TTCTAAACTTA  | CCTGTTGCCG<br>GGACAAACGGC | GTCTTGGCAT<br>CAGAACGCTA  |
| 1891 | GATTATCATA<br>CTAATAGTAT | TAATTTCGT<br>ATTAAGACAA   | TGAAATTACGT<br>ACTTAATGCA | TAAGGCATGTA<br>ATTGTACAT | ATAATTAAACA<br>TATTAAATTGT | TGTAATGCA<br>ACATTACGTA   | GACGTTATT<br>CTGCAATAAA   |

FIG.\_ 46D

FIGURE 46E

2591 CGTTCTTTAA TAGTGGACTC TTGTTCCAAA CTGGAAACAAC ACTCAACCCCT ATCTCGGTCT ATTCTTTGAA  
GCAAGAAATT ATCACCTGAG AACAAAGGTT GACCTTGTGG TGAGTGTGG TAGAGCCAGA TAGAAAACCT

2661 TTATTAAGGG ATTGTCGCGA TTTCGGCCTA TTGGTTAAAA AATGAGCTGA TTAACAAA ATTAAACGGG  
AAATATTCCC TAAACGGCT AACCAATT TTACTCGACT AAATGTGTTT TAAATTGCGC

2731 AATTTTAACA AAATTAAC GCTTACAATT TAGGTGGCAC TTTTCGGGA AATGTGGCG GAACCCCTAT  
TTAAATTTGT TTATAATTG CGAATGTTAA ATCCACCGTG AAAAGCCCT TTACACGGC CTGGGGGATA

2801 TTGTTTATT TTCTAAATAC ATTCAAATAT GTATCCGGTC ATGAGACAAT AACCCCTGATA ATGCTTCAA  
AACAAATAAA AAGATTATG TAAGTTATA CATAGGGAG TACTCTGTA TTGGGACTAT TTACGAAGTT

2871 TAATATGAA AAGGAAAGAG TATGAGTATT CAACATTTCG GTGTCGCCCT TATTCCCTTT TTGGGGCAT  
ATTATAACTT TTTCCTTCTC ATACTCATAA GTTGTAAAGG CACAGGGGA ATAAGGAAA AACGGCGTA

2941 TTGCCCCCTTC TGTGTTTGCT CACCCAGAAA CGCTGGTGA AGTAAAGAT GCTGAAGATC AGTTGGGTGC  
AACGGAAAGG ACAAAACGA GTGGGTCTT GCGACCCTT TCATTCTCTA CGACTCTCTAG TCAACCCACG

3011 ACGAGTGGGT TACATCGAAC TGGAATCTCAA CAGCGGTAAAG ATCCCTGAGA GTTTTGGCCC CGAAGAACGT  
TGCTCACCCA ATGTTAGCTTG ACTTAGAGTT GTGCCATT TAGGAACTCT CAAAGGGG GCTTCTTGCA

3081 TTCCCAATGAA TGAGCACTTT TAAAGTTCTG CTATGTTGCG CGGTATTATC CCGTATGAC GCGGGCAAG  
AAAGGTTACT ATCTGTGAAA ATTCAAGAC GATACACCGC GCCATAACTG CGGCCCGTTC

3151 AGCAACTCTGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTTGAGTAC TCACCACTCA CAGAAAGCA  
TGTTGAGCC AGCGGGCTAT GTGATAAGAG TCTTACTGAA CCAACTCATG AGTGGTCAGT GTCTTTTCGT

3221 TCTTACGGAT GGCATGACAG TAAGAGATT ATGAGCTGCT GCCATAACCA TGAGTGTATAA CACTGGGCC  
AGAATGCCTA CGTACTGTC ATTCTCTTAA TACGTACAGA CGGTATTGGT ACTCACTATT GTGACGCCGG

3291 AACTTACTTC TGACAACGGAT CGGAGGACCG AAGGAGCTAA CGGCTTTTT GCACAACTG GGGGATCATG  
TGTGAAATGAAAG ACTGTTGCTA GCCTCCTGGC TTCTCTGATT GGGAAAAA CGTGTGTAC CCCCTAGTAC

3361 TAACTGGCCT TGATCGTTGG GAAACGGAGG TGATGAAAC CATAACAAAC GACGGAGGCTG ACACCAAGAT  
ATTGAGCGGA ACTAGCAACC CTGGCCCTCG ACTACTTCG GTATGGTTG CTGCTCGCAC TGTGGTGTCA

|      |                                                      |                                             |
|------|------------------------------------------------------|---------------------------------------------|
| 3431 | GCCTGTAGCA ATGGCAACAA CGTTGGCAA ACTATTAAC            | GGCGAACTAC TTACTCTAGC TTCCCCGGCAA           |
|      | CGGACATCGT TACCGTTGTT GCAACGGTT TGATAATTGA           | CCGCTGTGATG AATGAGATCG AAGGGCCGTT           |
| 3501 | CAATTAAATAG ACTGGATGGA GGGGATAAA GTTGCAGGAC          | CACTTCTGCG CTCGGCCCTT CCGCTGTGATG           |
|      | CGGCCTATTG CAACGTCCGT GTGAAGACGC GAGCCGACCGA         | GGCCGACTCG                                  |
| 3571 | GGTTTATTGC TGATAAAATCT GGAGCGGGTG AGCGGTGGTC         | TCGGGGTATC ATTGCAGCAC TGGGGCCAGA            |
|      | CCTCGGCCAC TCGCACCCAG AGGCCATAG TAACGTGTG ACCCGGGTCT |                                             |
| 3641 | TGGTAAGCCC TCCCCTATCG TAGTTATCTA CACGACGGGG          | AGTCAGGCAA CTATGGATGA ACCGAATAGA            |
|      | ACCATTGGG AGGGCATAGC ATCAATAGAT GTGCTGCCCTC          | TCAGTCCCGTT GATAACCTACT TGCTTTATCT          |
| 3711 | CAGATCGCTG AGATAGGTGC CTCACTGATT AAGCATTGGT          | AACTGTCAGA CCAAGTTTAC TCATATATAC            |
|      | GTCTAGGCAC TCTATCCACG GAGTGACTAA TTGACAGTCT          | GGTCAAATG AGTATATATG                        |
| 3781 | TTTAGATTGA TTAAAAACTT CATTTTTAAT TTAAAGGAT           | CTAGGTGAAG ATCCCTTTTG ATAATCTCAT            |
|      | AAATCTAACT AAATTGAA GTAAAAATAA AATTTCCTA GATCCACTTC  | TAGAAAGAT GGTCAAATG ATTAGAGTA               |
| 3851 | GACCAAAATC CCTTAACGTG AGTTTTCGTT CCACTGAGCG          | TCAGACCCCCG TAGAAAGAT CAAAGGATCT            |
|      | CTGGTTTTAG GGAATTGCAC TCAAAAGCAA GGTGACTCGC          | AGTCTGGGGC ATCTTTCTA GTTTCCCTAGA            |
| 3921 | TCTTGTAGATC CTTTTTTCTC GGGCGTAATC TGCTGCTTGC         | AAACAAAAAA ACCACCGCTA CCAGGGTGG             |
|      | AGAACTCTAG GAAAAAAGA CGGGCATTAG ACGACGAACG           | TTTGTGTTTT TGTTGGGAT GGTGCTCTA              |
| 3991 | TTTGTGTTGCC GGATCAAGAG CTACCAACTC TTTTCCGAA          | GGTAACCTGGC TTCAAGAAGT CGCAGATAACC          |
|      | AAACAAACGG CCTAGTTCTC GATGGTTGAG AAAAAGGCTT          | CCATTGACCG AAGTTCTTGA AAGTCGTCTC GCGATGTATG |
| 4061 | AAATACTGTC CTCTCTAGTGT AGGCCACAC TTCAAGAACT          | CTGTAGCACC GCCTACATAC                       |
|      | TTTATGACAG GAAGATCACA TCGGCATCAA TCCGGTGGTG          | GACATCGTGG CGGATGTATG                       |
| 4131 | CTCGCTCTGC TAATCCTGTT ACCAGTGGCT GCTGCCAGTC          | GTGTCCTTAC GGGTTGGACT                       |
|      | GAGCGAGACG ATTAGGACAA TGGTCACCGA CGACGGTCAC          | CGCTATTCA CGCAGATGG CCCAACCTGA              |
| 4201 | CAAGACGATA GTTACCGGAT AAGGGCGAC GGTCGGGCTG           | AACGGGGGT TCGTGCACAC AGCCCCAGCTT            |
|      | GTTCTGCTAT CAATGGCCTA TTCCGGCTCG CCAGCCCCGAC         | AGCACGTGTG TCGGGTCCGAA                      |

FIG.-46G

4271 GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGGCCAC GCTTCCCGAA  
CCTCGCTTGC TGGATGTGGC TTGACTCTAT GGATGTGGCA CTCGATACTC TTTGCGGGTG CGAAGGGCTT

4341 GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTGC GAACAGGAGA GGGCACGAGG GAGCTTCCAG  
CCCTCTTTC GCCTGTCCAT AGGCCATTG CCGTCCCAGC CTTGTCCTCT CGCGTGTCTCC CTCGAAGGGTC

4411 GGGGAACGC CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCCCTCTGA CTTGAGCGTC GATTTTTGTG  
CCCCTTGTGC GACCATAGAA ATATCAGGAC AGCCCAAAGC GGTGGAGACT GAACTGGCAG CTAAAAACAC

4481 ATGCTCGTCA GGGGGGGGA GCCTATGGAA AACGCCAGC AACGCCAGC AACGCCAGC AACGCCAGC  
TACGAGCAGT CCCCCGGCT CGGATACCTT TTTGCGGTGC TTGGGCCGGA AAAATGCCAA GGACCGGAAA

4551 TGCTGCCCTT TTGCTCACT GTCTTTCCT GCGTTATCCC CTGATCTGT GGATAACCGT ATTACCGCCT  
ACGACCGAA AACGAGTGTAA CGAAAGGA CGCAATAGGG GACTAAGACA CCTATTGGCA TAATGGGGA

4621 TTGAGTGAGC TGATACCGCT CGCCGCAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA  
AACTCACTCG ACTATGGGA GCGGGCTCGG CTTGCTGGCT CGCGTGCCTC AGTCACCTCGC TCCTTCTGCCT

4691 AGAGGCCCA ATACGCAAC CGCCCTCTCCC CGGGCGTTGC CCGATTCATT AATGCAGCTG GCACGACAGG  
TCTCGGGGT TATGGCTTTCG GCGGAGAGGG GCGCGCAACC GGCTAAGTAA TTACGTGAC CGTGTGTGCC

4761 TTTCGGACT GGAAGGGGG CAGTGAGCGC AACGCAATTAA ATGTGAGTTA GCTCACTCAT TAGGCACCC  
AAAGGGCTGA CCTTTCGCCG GTCACTCGCG TTGGTTAAT TACACTCAAT CGAGTGAAGTA ATCCGTGGGG

4831 AGGCTTACA CTTTATGCTT CCGGCTCGTA TGTGTGAGC GGATAACAAAT TTCACACAGG  
TCCGAAATGT GAAATACGAA GGCGGAGCAT ACAACACACC TAAACACTCG CCTATTGTTA AAGTGTGTCC

4901 AACAGCTAT GACCATGATT AGGCCAACGG CGCAATTAAAC CCTCACTAA GGGAAACAAA GCTGG  
TTTGTGATA CTGGTACTAA TGCGGTTCGC GCGTTAATTG GGAGTGTGATT CCCTTGTGTTT CGACC

ECORI

BBSHII

~~~~~

FIG._46H

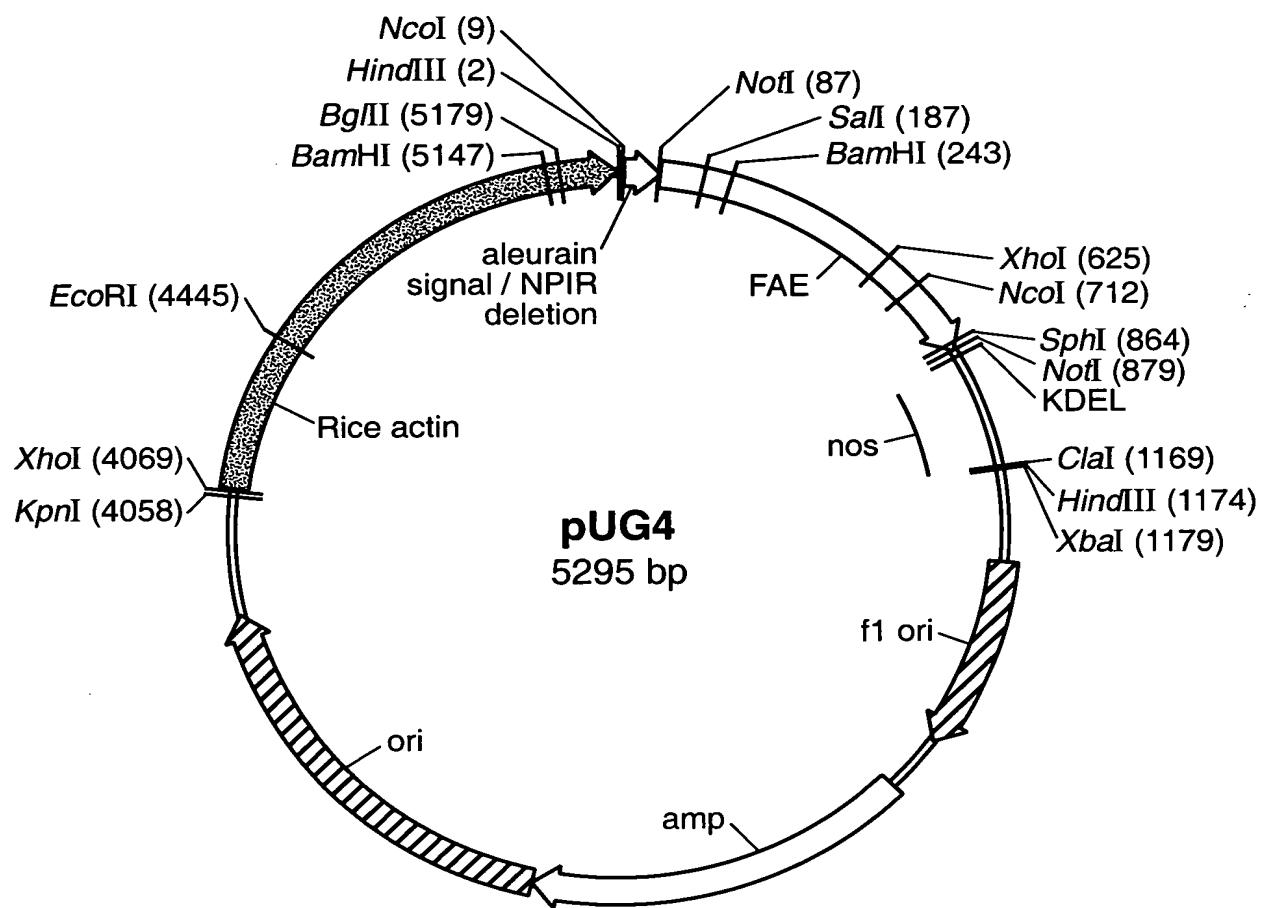


FIG._47A

NcoI

HindIII

1 M A H A R V L L A L A V L A T A A V A V
AAGCTTACCA TGGCCCACGGC CGCGTCTCTC CTCACGGCGC TCGCCGGC TCGCCGGCT GCCCACGGCC GCCGTGCCG

NotI

71 A S S R A A S T Q G I S E D L Y S R L V E M
TCGCCTCTC CGCGCGGCC GCCTCACGC AGGGCATCTC CGAAGACCTC TACAGCCGT TAGTCGAATT
Sall

141 A T I S Q A A Y A D L C N I P S T I I K G E K
GGCCACTATC TCCCAAGCTG CCTACGGCGA CCGTGTGCAAC ATTCCGTGCGA CTATTATCAA GGGAGAGAAA
BamHI

211 I Y N S Q T D I N G W I L R D D S S K E I I T V
ATTACAAATT CTCAAACTGA CATTAAACGGGA TGGATCCTCC GCGACGACAG CAGCAAGAA ATAACATACCG
F R G T G S D T N L Q L D T N Y T L T P F D T
281 TCTTCCGGGG CACTGGTAGT GATACTGAATC TACAACCTCGA TACTAACTAC ACCCTCAAGC CTTTCGACAC
L P Q C N G C E V H G G Y Y I G W V S V Q D Q
351 CCTACCAAA TGCAACGGTT GTGAAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCGT CCAGGACCAA
V E S L V K Q Q V S Q Y P D Y A L T V T G H X L
421 GTCGAGTCGC TTGTCAAAACA GCAGGTTAGC CAGTATCCGG ACTACGGGT GACCGTGACC GGGCAACKCC
G A S L A A L T A A Q L S A T Y D N I R L Y T
491 TCGCGGCCCTC CCTGGCGGCC CTCACGTGCG CCCAGCTGTC TGCGACATAAC GACAACATCC GCCCTGTACAC
XbaI

561 F G E P R S G N Q A F A S Y M N D A F Q A S S
CTTCGGCGAA CGCGCGAGGG GCAATCGAGG CTTCGCGTCG TACATGAACG ATGCCTTCGA AGCCTCGAGGC
P D T T Q Y F R V T H A N D G I P N L P P V E Q
631 CCAGATACGA CGCAGTATT CCGGGTCACT CATGCCAACG ACGGCATCCC AACCTGCC CCGGTGGAGC
NcoI

FIG.-47B

• G Y A H G G V E Y W S V D P Y S A Q N T F V C .
701 AGGGGTACGC CCATGGGGGT GTAGAGTACT GGAGGGTTGA TCCTTACAGC GCCCAGAACCA CATTGGTCTG
• T G D E V Q C C E A Q G G Q G V N A H T T Y
771 CACTGGGGAT GAAGTGGCAGT GCTGTGAGGC CCAGGGGGGA CAGGGTGTGA ATAATGGCA CACGACTTAT

SphI

F G M T S G A C T W P V A A E P L K D E L *
841 TTTGGGATGA CGAGGGGGC ATGACACTGG CCGGGTGGG CCGGGGAACC ACTGAAGGAT GAGGCTGTAAA
911 GAAGCAGATC GTTCAAAAT ATGGCAATA AGTTCTTAA GATTGAATCC TGTTGCCGGT CTGCGATGAA
981 TTATCATATA ATTCTCTGTG AATTACGTTA AGCATGTAAT AATTACATG TAATGCATGA CGTTATTATAT
1051 GAGATGGGT TTTATGATTA GAGTCCCCGA ATTATACATT TAATACGGGA TAGAAACAA ATAATAGGCC
NotI

HindIII

Clal

~~~~~  
1121 GCAAACTAGG ATAAATTATC GCGGGGGTG TCATCTATGT TACTAGATCG ATAAGCTCT AGAGGGCCCG  
1191 GTGGAGCTCC AATTGCCCT ATAGTGAGTC GTATACGCG CGCTCACTGG CGTCGTGTTT ACAACGTCGT  
1261 GACTGGAAA ACCCTGGCGT TACCCAACT TACCTGGCTG CAGCACATCC CCCTTTCGCG AGCTGGCGTA  
1331 ATAGCGAAGA GCCCCGACC GATCGCCCT CCCAACAGTT GCGCAGCCTG AATGGCGAAT GGACGGCGCC  
1401 CTGTAGGGC GCATTAAGCG CGGCGGGGT GGTGGTTACG CGCAGCGTGA CGCGTACACT TGCCAGCGCC  
1471 CTAGGCCCG CTCCCTTCGC TTCTCTCCCT TCCTTTCTCG CCACGTTGC CGGCTTTCCC CGTCAAGCTC  
1541 TAAATCGGG GCTCCCTTTA GGGTTCCGAT TTAGTGCTT ACGGCACCTC GACCCCCAAA AACTTGATTA  
1611 GGGTGATGGT TCACGTAGTG GGCATCGC CTGATAGACG GTTTTCGCC CTTTGACGT GGAGTCCACG  
1681 TTCTTTAATA GTGGACTCT GTTCCAAACT GAAACAAAC TCAACCCAT CTCGGCTAT TCCTTTGATT  
1751 TATAAGGGAT TTGCGCGATT TCGGCCATT GGTAAAGAA TGAGCTGATT TAACAAAT TAAAGGGAA  
1821 TTTAACAAA ATATTAACGC TTACAATTAA GTGCGCACTT TTCGGGGAAA TGTGCGCGGA ACCCCTATT  
1891 GTTTATTT CTAATACAT TCAAATATGT ATCCGCTCAT GAGACAATAA CCCTGATAAA TGCTTCATAA  
1961 ATATTGAAA AGGAAGAGTA TGAGTATTCA ACATTTCCGT GTCGGCCCTTA TCCCTTTTG TGCGGATT  
2031 TGCCCTTCCTG TTTTGCTCA CCCAGAACG CTGGTGAAAG TAAAGATGC TGAAGATCAG TTGGGTGCAC  
2101 GAGTGGGTTA CATCGAACGT GATCTAACCA GCGGTAAAGAT CCTGAGGT TTTCGCCCG AAGAACGTTT  
2171 TCCAATGATG AGCACTTTA AAGTTCTGCT ATGTTGGCGC GTATTATCCC GTATTGACGC CGGGCAAGAG  
2241 CAACTGGTC GCGGCATACA CTATTCTCAG AATGACTTGG TTGAGTACTC ACCAGTCACA GAAAGCATT  
2311 TTACGGATGG CATGACAGTA AGAGAAATTAT GCAGTGCTGC CATAACCATG AGTGATAACA CTGGGGCCAA  
2381 CTTACTCTG ACAACGATCG GAGGACCGAA GGAGCTAACC GCTTTTTGC ACAACATGGG GGATCATGTA  
2451 ACTCGGCCTTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA TACCAAACGA CGAGGTTGAC ACCACGATGC  
2521 CTGTAGCAAT GGCAACAAACG TTGGCAAC TATTAACG CGAACTACTT ACTCTAGCTT CCCGGCAACA

FIG.-47C

|                       |             |             |            |             |            |             |             |
|-----------------------|-------------|-------------|------------|-------------|------------|-------------|-------------|
| 2591                  | ATTAATAGAC  | TCGGATGGGG  | CGGATAAAGT | TGCAGGGACCA | CTTCTGGGCT | CGGCCCTTCC  | GGCTGGCTGG  |
| 2661                  | TTTATTGCTG  | ATAAATCTGG  | AGCCGGTGA  | CGTGGGTCTC  | GGGGTATCAT | TGCAGGCACTG | GGGCCAGATG  |
| 2731                  | GTAAGCCCTC  | CGGTATCGTA  | GTTATCTACA | CGACGGGGAG  | TCAGGCAACT | ATGGATGAAAC | GAAATAGACAA |
| 2801                  | GATCGGTGAG  | ATAGGTGCTCT | CACTGATTAA | GCATTTGGTA  | CTGTCAGACC | AGTTTACTC   | ATATAACTT   |
| 2871                  | TAGATTGATT  | TAAAACTTCA  | TTTTTAATT  | AAAAGGATCT  | AGGTGAAGAT | CCTTTTGTAT  | AATCTCATGA  |
| 2941                  | CCAAAATCCC  | TAAACGTGAG  | TTAACGTTC  | ACTGAGGCTC  | AGACCCCGTA | GAAAAGATCA  | AAGGATCTTC  |
| 3011                  | TTGAGATCCCT | TTTTTTCTGC  | GGGTAATCTG | CTGCTTGCAA  | ACAAAAAAAC | CACCGCTAAC  | AGGGGTGGTT  |
| 3081                  | TGTTTGCAGG  | ATCAAGAGCT  | ACCAACTCTT | TTTCCGAAGG  | TAACTGGCTT | CAGCAGAGGG  | CAGATAACAA  |
| 3151                  | ATACTGTCT   | TCTAGTGTAG  | CCGTAGTTAG | GCCACCCACTT | CAAGAACTCT | GTAGGCACGCC | CTACATACCT  |
| 3221                  | CGCTCTGCTA  | ATCCTGTTAC  | CAGTGGCTGC | TGCCAGTGGC  | GATAAGTCGT | GTCTTACCGG  | GTGGAACCTCA |
| 3291                  | AGACGATAGT  | TACCGGATAA  | GGGCCAGGG  | TCGGGCTGAA  | CGGGGGTTTC | GTGCACACAG  | CCAGGCTTGG  |
| 3361                  | AGCGAACGAC  | CTACACCGAA  | CTGAGATAAC | TACAGCCTGA  | GCTATGAGAA | AGGCCACCGC  | TTTCCCGAAGG |
| 3431                  | GAGAAAGGGCG | GACAGGTATC  | CGGTAAGGGG | CAGGTCTCGGA | ACAGGAGGGC | GCACGAGGG   | GCTTCCAGGG  |
| 3501                  | GGAAACGCCT  | GGTATCTTTA  | TAGTCCTGTC | GGGTTTCGCC  | ACCTCTGACT | TGAGGCTGCA  | TTTTTGTGAT  |
| 3571                  | GCTCGTCAGG  | GGGGGGAGC   | CTATGGAAA  | ACGCCAGCAA  | CGGGCCCTTT | TAAGGGTCTCC | TGGCCTTTTG  |
| 3641                  | CTGGCCCTTT  | GCTCACATGT  | TCTTTCCTGC | GTATCCCCCT  | GATTCTGTGG | ATAACCGTAT  | TACCGCCTT   |
| 3711                  | GAGTGAGCTG  | ATACCGCTCG  | CCGCAGCCGA | ACGACCCGAGC | GCAGCGAGTC | AGTGAGCGAG  | GAAGGGAAAG  |
| 3781                  | AGCGCCAAT   | ACGCAAACCG  | CCTCTCCCCG | CGCGTTGGCC  | GATTCACTAA | TGAGGTGGC   | ACGACAGGTT  |
| 3851                  | TCCCGACTGG  | AAAGGGGCA   | GTGAGGCCA  | CGCAATTAA   | GTGAGTTAGC | TCACTCATTAA | GGCACCCCCAG |
| 3921                  | GCTTTACACT  | TATGCTTCC   | GGCTCGTATG | TTGTGTGGAA  | TTGTGTGGGG | ATAACAATT   | CACACAGGAA  |
| 3991                  |             |             |            |             |            |             |             |
| ACAGCTATGA CCATGATTAC |             |             |            |             |            |             |             |
| GCCAAGGGCG            |             |             |            |             |            |             |             |
| CAATAACCC             |             |             |            |             |            |             |             |
| TCACTAAAGG            |             |             |            |             |            |             |             |
| GAACAAAGC             |             |             |            |             |            |             |             |
| TGGGTACCGG            |             |             |            |             |            |             |             |
| XbaI                  |             |             |            |             |            |             |             |
| 4061                  | GCCCCCCTC   | GAGGTCACTTC | ATATGCTTA  | GAAGAGAGTC  | GGGATAAGTC | AAAATAAAAC  | AAAGGTAAGA  |
| 4131                  | TTACCTGGTC  | AAAAGTGA    | AAATCACTTA | AAAGGTGGTA  | TAAGTAAAT  | ATCGGTTATA  | AAAGGTGGCC  |
| 4201                  | CAAAGTGA    | AAATTCCTT   | TCTACTATTA | TAATAATTGA  | GGATGTTTTG | TGGGTACTTT  | GATACGTCTAT |
| 4271                  | TTTGTATGA   | ATTGGTTTT   | AAGTTTATTC | GCGATTGGAA  | AATGCATATC | TGTATTGAG   | TGGTTTTA    |
| 4341                  | AGTCGTTGC   | TTTTGTAAAT  | ACAGAGGGAT | TTGTATAAGA  | AATATCTTTA | AAAACCCAT   | ATGCTTAATT  |
| KpnI                  |             |             |            |             |            |             |             |
| 4411                  | GACATAATT   | TGAGAA      | AAA        | TATATATTCA  | GGCGAATTCC | ACAATGAAACA | ATAATAAGAT  |
| 4481                  | TGCCCGCGTT  | GCAGCGATGG  | GTATT      | AAAGTAAATA  | TAGTAAATA  | AAAGATAAAAC | TAGACTCAA   |
| 4551                  | AAAACACCC   | CTAAAGTCCT  | AAAGCCCAA  | GTGCTATGCA  | CGATCCATAG | CAAGCCAGC   | CCAACCCAAAC |
| 4621                  | CCAAACCAAC  | CCACCCCGT   | GCAGCCAACT | GGCAAATAGT  | CTCCACCCCC | GGCACTATCA  | CCGTGAGTTG  |
| EcoRI                 |             |             |            |             |            |             |             |

**FIG.\_47D**

|      |             |            |             |            |             |            |            |
|------|-------------|------------|-------------|------------|-------------|------------|------------|
| 4691 | TCCGCACCCAC | GGCACGTCTC | GCAGCCAAA   | AAAAAAAG   | AAAGAAAAAA  | AGAAAAAGA  | AAACAGCAG  |
| 4761 | GTGGGTCCGG  | GTGGTGGGG  | CGGGAAAAGC  | GAGGAGGATC | GCGAGCAGCG  | ACGAGGCCCG | GCCTCCCTC  |
| 4831 | CGCTTCACAA  | GAAACGGCCC | CCATCGCCAC  | TATATACATA | CCCCCCCCTC  | TCCTCCCATC | CCCCCAACCC |
| 4901 | TACCAACCACC | ACCACCACCA | CCTCCCTCCCC | CCTCGCTGCC | GGACGAGCGAG | CTCCTCCCCC | CTCCCCCTCC |
| 4971 | GCCGCCGCCG  | GTAACCACCC | CGCCCCCTCTC | CTCTTCTCTT | CTCCGTTTTT  | TTTTTCTGCT | GGTCTCGAT  |
| 5041 | CTTGGCCCTT  | GGTAGTTGG  | GTGGCGGAGA  | GGGGCTTCGT | CGCCCAAGATC | GTTGCGGGGG | AGGGGGGGGA |

BamHI

~~~~~

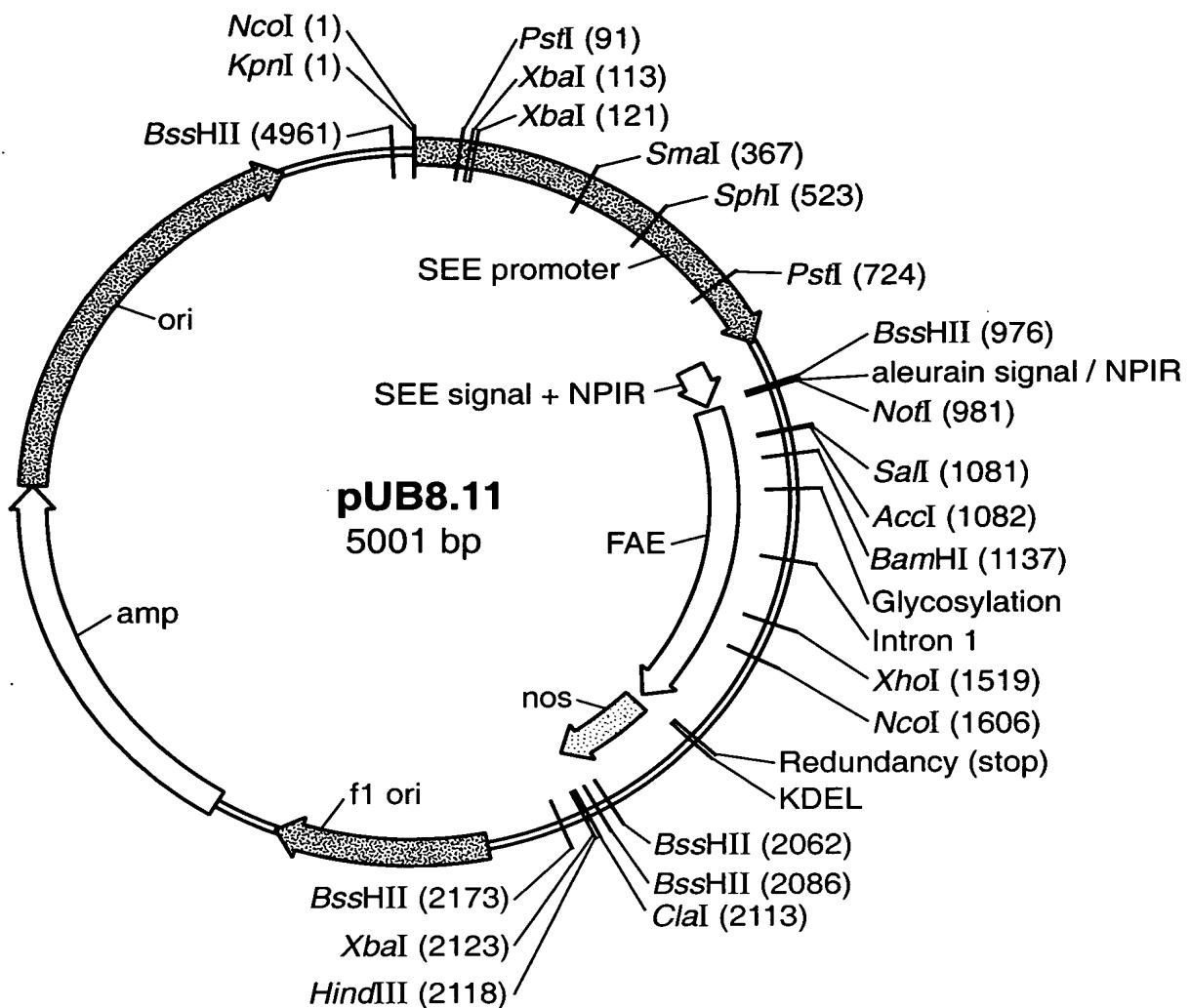
| | | | | | | | |
|------|-----------|------------|-------------|-----------|----------|------------|------------|
| 5111 | TCTCGGGCT | GGCGTCTCCG | GGCGTGAAGTC | GGCCGGATC | CTCGGGGA | ATGGGGCTCT | CGGATGTAGA |
|------|-----------|------------|-------------|-----------|----------|------------|------------|

BgIII

~~~

|      |             |            |            |            |              |            |           |
|------|-------------|------------|------------|------------|--------------|------------|-----------|
| 5181 | TCTCTTCTCTT | TCTCTTCTT  | TGTGGTAGAA | TTTGAATCCC | TCAGGCATTTGT | TCATCGGTAG | TTTTCTTTT |
| 5251 | CATGATTTGT  | GACAAATGCA | GCCTCGTGC  | GAGCTTTTTT | GTAGC        |            |           |

**FIG.\_47E**



**FIG.\_48A**

NcoI

~~~~~  
KpnI

1 CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATGGCT ACAATATCTG
GTACCCGGTC CATATTATAA CCCATAGAG TTCGTTATT AGCTTATAG TGGTAACCGA TGTATAGAC

PstI

~~~~~

71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC  
TCGAGGGCTCA AGACTGACGT CAGACCTACT GCGACAAACA TAGATCTGA GATCTATCGT GTGGGTGTCG

141 ACCTACAGGA GTGCGGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCCTAC CTCCGTGAGT  
TGGATGTCTT CACGCTGTGA ACACCTGACA TCATCACAAAC CTCTGCTCG AGAAAGGATG GAGGACTGCA

211 TGCCTGGCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGGCTC CCAACAAAT ATCGTCCCC  
ACGGCGCAA CAGGTAAGGT TGCCGTAGTG AGAGTTGGT AGTGGCGAG GTTGTTTA TAGCAGGGGG

281 ATGTCTGGC GGAGAGAGAG TACATACATG CTGTCGGGCC GTTTTGCT GAATCTCGCT TCCACTGGCC  
TACAGAACCG CCTCTCTCTC ATGTATGTAC GACAGCGCGG CAAACACAGA CTTAGAGCGA AGGTGACCGG

SmaI

351 AATCAGCTCA GCTCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA  
TTAGTCGAGT CGAGGGCCCT CGAGTGTAGTA AGTTCTAGGG TAGCAGCAGC AGTGGGACCC GCAGTACCT

421 TGGAAAGAA CCTCCGTGTC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAT  
ACCTTTCTCTT GGAGGCAACG AGCCTACTCA GTCCGTATAG GGGCTGTCT CATGACGTTTC TATTGGGTAA

SphI

491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTTCGGG TTTTGTGG CTTAATTGAC TTATTTTG  
AGTCTAAGGG GGTTATCTCT TTCAATATCGT ACGAAAAGCCC AAAACAAACC GAATTAACG AAATAAAAC

561 TTGGAGTTGA ATGCTGATTT GTGCTGTAA ATGCCCAACC ATCTGAATAT CGAGACGGAT ATAGGGCTGG  
AACCTCAACT TACGACTAAA CAACACATT TACGGGTGG TAGACTTATA GCTCTGCCTA TTATCCGACC

**FIG.- 48B**

631 CTAATTAAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTCT GGGCATTACA GCTGGAGGCT  
GATTAATAA ATATCGTTCT AAGACATCAC GTGTAGCGTT TATAGAAAGA CCCGTAATGT CGACCTCCGA

PstI

701 TCATCAGCCCT GAAACACTCT GCAGAGCCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGATAAAAC  
AGTAGTCGGA CTTTGTGAGA CGTCTCGGAC TTCGTCACC ACTTCGCAAC GCTACTCTAC CCATATTTTG

771 CCCGGCACCC GGGACGGAG CTCGGCCCTA CCAGTACCAT CTCGCCTCGC TCCCGCTGCC GGACGACCCA  
GGGGCCGTGG CCCTGGCTC GAGGGGGAT GGTCATGGTA GAGGGGAGCG AGGGGAGGG CCTTGCTGGGT

841 GTAAAAATACT GTTGGCCCACT CGCCGGCGAG ATGGCCCACG GCCGCATCCT CTCTTGGCG CTCGCCGTCT  
CATTTTTATGA CAAACGGGTGA CGGGCCGCTC TACCGGTGCG CGGGGTAGGA GAGAACCC GAGGGCAGA

BssHII

911 TGGCCACCGC CGGGTGGCC GCCGCATCNT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGGGCGC  
ACCGGGTGGCG GGGCCACCGG CGGGTAGNA ACCGGCTGAG GTTGGGTAG GCGGGCAGT GGCTCGGGCG  
Not I

981 GGCCGCCTCC ACGCAGGGCA TCTCCGAAGA CCTCTACAGC CGTTTAGTCG AAATGGCCAC TATCTCCCAA  
CGGGCGAGG TGCGTCCCGT AGAGGCTTCT GGAGATGTGCG GCAAATCAGC TTACCGGTG ATAGGGGT  
SalI

1051 GCTGCCTACG CGGACCTGTG CAACATTCCG TCGACTTATA TCAAGGGAGA GAAAATTAC AATTCTCAA  
CGACGGATGC GGCTGGACAC GTTGTAAAGGC AGCTGATAAT AGTTCCCTCT CTTTTAAATG TTAAGAGTT

FIG.\_48C

140 / 154

BamHI

1121 CTCGACATAA CGGATGGATC CTCCGGGACCG ACAGGAGCAA AGAAATAATC ACCGTTCTTCC GTGGCACTGG  
 GACTGTAATT GCCTACCTAG GAGGGCTGTC TGTGTCGTT TCTTTATTAG TGGCAGAAGG CACCGTGAC

1191 TAGTGTACG AATCTACAAAC TCGATACTAA CTACACCCCTC ACGCCTTTCG ACACCCTAAC ACAATGCAAC  
 ATCACTATGC TTAGATGTT AGCTATGATT GATGTGGGAG TGGGAAAGC TGTGGGATGG TGTACGTTG

1261 GGTGTGAAG TACACGGTGG ATATTATTT GGATGGTCT CGGTCCAGGA CCAAGTCCAG GTCGCTTGTCA  
 CCAAACACTC ATGGGCCACC TATAATAAA CCTACCCAGA GGCAGGGCCT GGTTCAGCTC AGCGAACAGT

1331 AACAGGAGGT TAGCCAGTAT CCGGACTTACCG CGCTGACCGT GACCGGCCAC KCCCTCGGGC CCTCCCTGGC  
 TTGTCGTCCTA ATCGGTCTATA GGCGACTGGCA CTGGCCGGTG MGGGAGCCGC GGAGGGACCG

1401 GGGCACTCACT GCGGCCAGGC TGTCCTGCAGC ATACGACAAC ATCCGGCTGT ACACCTTCGG CGAACCGCGC  
 CCGTGAGTGA CGGGGGTGTG ACAGACGCTG TATGCTGTG TAGGGGACA TGTGGAAAGCC GCTTGGCGC

XbaI

1471 AGGGCAATC AGGGCTTTCGC GTCGTACATG AACGATGGCT TCCAAGGCTC GAGGCCAGAT ACGACGGCAGT  
 TCGCCGTAG TCGGAAGCG CAGCATGTAC TTGCTACGGA AGGTTGGAG CTCGGTCTA TGCTGGTCA  
 ~~~~~ NcoI ~~~~~

1541 ATTTCCGGGT CACTCATGCC AACGACGGCA TCCCAAACCT GCCCCGGTG GAGCAGGGGT ACGCCCATGG
 TAAAGGCCA GTGAGTACGG TTGCTGCCGT AGGGTTGGAA CGGGGGCCTAC CTCGTCCTA TGCGGTAC

1611 CGGTGTAGAG TACTGGAGC TTGATCCTTA CAGGCCAG AACACATTG TCTGCACTGG GGATGAAGTG
 GCCACATCTC ATGACCTCGC AACTAGGAAT GTCGGGGTC TTGTGTAAAC AGACGTGACC CCTACTTCAC

1681 CAGTGCTGTG AGGCCAGGG CGAACAGGGT GTGAATAATG CGCACACGAC TTATTTGGG ATGACGAGCG
 GTCACGACAC TCGGGTCCC GCCTGTCTCA CACTTATTAC GGTGTGCTG ATAAAAACCC TACTGCTCGC

1751 GAGCCTGTAC ATGGTGTACA GTCATTTCA CGTACCCGAG TGTACCAAGA AAGATGGATG TCCCTGGAGAG
 CTCGGACATG TACCACTAGT CAGTAAAGTC GGAGGGCTC ACATGGTCTT ACACCTAC AGGACCTCTC

FIG.-48D

| | | |
|------|---|--------|
| 1821 | GGGGCCGGT AACCACTGAA GGATGAGCTG TAAAGAAGCA GATCGTCAA ACATTTGCCA ATAAAGTTTC CCCCGGGCA TGGTGACTT CCTACTCGAC ATTCTTCGT CTAGCAAGT TATTCAAAAG | BssHII |
| 1891 | TTAAGATTGA ATCCCTGTTGC CGGTCTTGCG ATGATTATCA TATAATTCT GTTGAATTAC GTTAAGGCATG AATTCTAACT TAGGACAACG GCCAGAACGC TACTAATAGT ATATAAAGA CAACTTAATG CATTCTGTC | XbaI |
| 1961 | TAATAATTAA CATGTAATGC ATGACGTTAT TTATGAGATG GGTTTTATG ATTAGAGTCC CGCAATTATA ATTATAATT GTACATTACG TACTGCAATA AATACTCTAC CAAAAATAAC TAATCTCAGG GCGTTAATAAT | BssHII |
| 2031 | CATTAAATAC GCGATAGAAA ACAAAATATA GCGGCCAAC TAGGATAAAT TATCGCGCGC GGTGTCATCT GTAAATTATG CGCTATCTT TGTTTTATAAT CGCGCGTTTG ATCCTATTTA ATAGCGCGC CCACAGTAGA | BssHII |
| 2101 | ATGTTACTAG ATCGATAAGC TTCTAGAGCG GCCGGTGGAG CTCCAATTCTG CCCTATAGTG AGTCGTATTA TACAATGATC TAGCTATTACG AAGATCTCGC CGGCCACCTC GAGGTTAAGC GGGATATCAC TCAGGATAATT | BssHII |
| 2171 | CGCGCGCTCA CTGGCCGTCG TTTACAACG TCGTGAATGG GAAAACCTG GCGTTACCCCA ACTTAATCGC GGCGGCCAGT GACCGGCAGC AAAATGTTGC AGCACTGACCC CTTTTGGGAC CGCAATGGGT TGAATTAGCG | BssHII |
| 2241 | CTTGGCAGCAC ATCCCCCTTT CGCAGCTGG CGTAAATAGCG AAGAGGGCCCG CACCGATCGC CCTTCCCAAC GAACGTCGTG TAGGGGAAA GCGTCGACCC GCATTATCGC TTCTCCGGGC GTGGCTAGCG GGAAGGGTTG | BssHII |
| 2311 | AGTTGCCAG CCTGAATGGC GAATGGGACCG CGCCCTGTAG CGGGGCATTA AGCGGGGGGG GTGTGGTGGT TCAACGGCTC GGACTTACCG CTTACCCCTGC GCGGACATC GCCGCTTAAT TCAGGCCGCC CACACCCA | BssHII |
| 2381 | TACGGCAGC GTGACCGCTA CACTTGCAG CGCCCTAGCG CCCGCTCCCTT TCGCTTTCTT CCCTTCCTT ATGGCGGTGC CACTGGGAT GTGAACGGTC GCGGATCGC GGGGAAAGAA AGCGGAGGAA GGGAAAGGAA | BssHII |

FIG._48E

2451 CTCGCCACCGT TCGCCGGCTT TCCCGTCAA GCTCTAAATC GGGGGCTCCC TTAGGGTTC CGATTTAGTG
GAGGGGTGCA AGGGCCGAA AGGGCAGTT CGAGATTAG CCCCCGAGGG AAATCCCAA GCTAAATC
GAAATGCCGT GGAGCTGGG TTTTTGAAC TAATCCACT ACCAAGTGCA TCACCCGGTA GCGGGACTAT
2591 GACGGTTTT CGCCCGACCC AAAAAACTTG ATTAGGGTGA TGGTTCACCGT AGTGGGCCAT CGCCCTGATA
CTGCCAAAGA GCGGAAACT GCAACCTCAG GTGCAAGAAA TTATCACCTG AGAACAAAGGT TTGACCTTGT
2661 ACACTCAACC CTATCTCGGT CTATCTTTT GATTATAAG GGATTTGCC GATTTCGGCC TATTTGGTTAA
TGTGAGTGG GATAGAGCCA GATAAGAAA CTAATATTG CCTAAAACGG CTAAGCCGG ATAACCAATT
2731 AAAATGAGCT GATTAAACAA AAATTAAACG CGAATTAA CAAATATTAA ACGCTTACAA TTAGGTGGC
TTTTACTCGA CTAAATTGTT TTTAAATTGC GCTTAAATT GTTTATAAT TGCGAATGTT AAATCCACCG
2801 ACTTTTCGGG GAAATGTGCG CGGAACCCCT ATTTGGTTAT TTTCTAAAT ACATTCAAT ATGTTATCCGC
TGAAAGCCC CTTTACACGC GCCTTGGGA TAAACAAATA AAAAGATTAA TGTAAGTTA TACATAGGGC
2871 TCATGAGACA ATAAACCTGA TAAATGCTTC AATAATATTG AAAAGGAAG AGTATGAGTA TCAACATT
AGTACTCTGT TATTGGACT ATTACGAAG TTATTATAAC TTTTCTTC TCATACTCAT AAGTTGTAAA
2941 CCGGTGCGC CTATTCCTT TTTTGGGC ATTTGGCTT CCTGTTTTG CTCACCCAGA AACGCTGGTG
GGCACAGCGG GAATAAGGA AAAAACGCCG TAAACGGAA GGACAAAAAC GAGTGGGTCT TTGCGACCA
3011 AAAGTAAAG ATGCTGAAGA TCAGTTGGGT GCACGAGTGG GTTACATCGA ACTGGATCTC AACAGCGGT
TTTCATTTC TAGACTTCT AGTCAACCCA CGTGCTCACC CAATGTAGCT TGACCTAGAG TTGTCGCCAT
3081 AGATCCTGTA GAGTTTCGC CCCGAAGAAC GTTTCCAAT GATGAGCACT TTAAAGTTC TGCTATGTGG
TCTAGGAACCTCATAAGCG GGGCTCTTG CAAAGGTTA CTACTCGTGA AAATTCAAG ACGATAACACC
3151 CGCGGTATTA TCCCGTATG ACGCCGGGCA AGAGCAACTC GGTGCGCGCA TACACTATT TCAGAAATGAC
GGGCCATAAT AGGGCATAAC TGCGGCCGT TCTCGTTGAG CCAGGGCGT ATGTGATAAG AGTCTTACTG
3221 TTGGTTGAGT ACTCACCAGT CACAGAAAG CATCTTACGG ATGGCATGAC AGTAAGAGAA TTATGGCAGTG
AACCAACTCA TGAGTGGTCA GTGTCTTTG TACCGTACTG TCATTCTCTT AATACGTCAC

3291 CTGCCATAAC CATGAGTGT AACACTGGG CCAACTTAAC TCTGACAACG ATCGGAGGAC CGAAGGAGCT
GACGGTATTG GTACTCACTA TTGIGACGCC GGTGAAATGA AGACTGTTGC TAGCCTCCCTG GCTCCTCGA

3361 AACCGCTTT TTGACACAACA TGGGGATCA TGTAACCTCA TGTAACCTCGC CTTGATCGT GGGAAACCGGA GCTGAATGAA
TTGGCGAAA AACGTTGTTGT ACCCCCTAGT ACATTGAGCC GAACCTAGCAA CCCTTGGCT CGACTTACTT

3431 GCCATACCAA ACGACGAGCG TGACACCACG ATGCCGTAG CAATGGCAAC AACGTTGCCG AACTATTAA
CGGTATGGTT TGCTGCTCGC ACTGTGGTGC TACGGACATC GTTACCGTTG TTGCAACCGCG TTGATAATT

3501 CTGGCGAACT ACTTAACTCTA GCTTCCCAGG AACAAATTAAT AGACTGGATG GAGGGGATA AAGTTGCAGG
GACCCTGTTGA TGAATGAGAT CGAAGGGCCG TTGTTAATTA TCTGACCTAC CTCGGCCTAT TTCAACGTC

3571 ACCACTTCTG CGCTCGGGCC TTCCGGCTGG CTGGTTTATT GCTGATAAAAT CTGGAGGCCG TGAGCGTGGG
TGGTGAAGAC CGAGGCCGG AAGGCCGACC GACAAATAA CGACTATTAA GACCTCGGCC ACTGGCACCC

3641 TCTCGGGTA TCATTGCAAC ACTGGGGCCA GATGGTAAGC CCTCCCGTAT CGTAGTTATC TACACGACGG
AGAGGCCAT AGTAACGTCG TGACCCCGGT CTACCATTCG GGAGGGCATA GCATCAATAG ATGTTGCTGCC

3711 GGAGTCAGGC AACTATGGAT AACCGAAATA GACAGATCGC TGAGATAGGT GCCTCACTGA TTAAGCATTG
CCTCAGTCGG TTGATAACCTA CTTGCTTTAT CTGTCATAGCG ACTCTATCCA CGGAGGTGACT AATTGTAAC

3781 GTAACTGTCA GACCAAGTT ACTCATATAT ACTTTAGATT GATTAAAAAC TTCATTTTA ATTAAAGG
CATTGACAGT CTGGTTCAAA TGAGTATATA TGAAATCTAA CTAATTTG AAGTAAAAAT TAAATTTC

3851 ATCTAGGTGA AGATCCTTT TGATAATCTC ATGACCAAA TCCCTTAACG TGAGTTTTG TTCCACTGAG
TAGATCCACT TCTAGGAAA ACTATTAGAG TACTGGTTT AGGGAAATTGC ACTCAAAAGC AAGGTGACTC

3921 CGTCAGACCC CGTAGAAAAG ATCAAAGGAT CTTCTTGAGA TCCTTTTTT CTGCGCGTAA TCTGCTGCTT
GCAGTCTGGG GCATCTTTTC TAGTTTCCCTA GAAGAACTCT AGGGAAAAAA GACGGCATT AGACGACGA

3991 GCAAAACAAAA AAACCAACGC TACCCAGGGT GGTTTGTGTTG CGGGATCAAG AGCTACCAAC TCTTTTCCG
CGTTTGTTT TTGGTGGCG ATGGTCGCC CAAACAAAC GGCCCTAGTTG TGATGGTT AGAAAAGGC

4061 AAGGTAACGT GCTTCAGCAG AGGGCAGATA CCAAATAACTG TCCTTCTAGT GTAGGCCGTAG TTAGGCCAC
TTCCATTGAC CGAAGTCGTC TCGCGTCTAT GGTTTATGAC AGGAAGATCA CATCGGCATC AATCCGGTGG

| | |
|------|---|
| 4131 | ACTTCAGAA CTCGTAGCA CGGCCTACAT ACCTCGCTCT GCTAATCCTG TTACCAAGTGG CTGCTGCCAG TGAAGTCTT GAGACATGTT GGGGATGTA TGGAGCGAGA CGATTAGGAC ATGGTCACC GACGACGGTC |
| 4201 | TGGCGATAAG TCGTGTCTTA CCGGGTTGGA CTCAGACGA TAGTTACCGG ATAAGGCCA GCGGTGGGC ACCGCTATTG AGCACAGAAT GGGCCAAACCT GAGTTCTGCT ATCAATGGCC TATTCCGGGT CGCCAGGCCG |
| 4271 | TGAACGGGGG GTTCGTGCAAC ACAGCCCCAGC TTGGAGCGAA CGACCTACAC CGAACCTGAGA TACCTACAGC ACTTGCCCCC CAAGCACGTG TGTGGGTG AACCTCGCTT GCTGGATGTG GCTTGACTCT ATGGATGTG |
| 4341 | GTGAGCTATG AGAAAGGGCC ACGGCTTCGGG AAGGGAGAAA GGCGGACAGG TATCCGGTAA GCGGCAGGGT CACTCGATACT TCTTCGCGG TGGGAAGGGC TTCCCTCTTT CGGCCTGTCC ATAGGCCATT CGCCGTCCTCA |
| 4411 | CGGAACAGGA GAGGCCACGA GGGAGCTTCC AGGGGAAAC GCCTGGTATC TTTATAGTCC TGTGGGGTT GCCTTGTCCT CTCGGTGTCT CCCTCGAAGG TCCCCCTTTG CGGACCATAG AAATATCAGG ACAGCCAAA |
| 4481 | CGCCACCTCT GACTTGAGCG TGATTTTGT TGATGCTCGT CAGGGGGCG GAGCCTATGG AAAAACGCCA CGGGTGGAGA CTGAAACTCGC AGCTAAAC ACTACGAGCA GTCCCCCGC CTGGATACCT TTGCGGGT |
| 4551 | GCAACGGGC CTTTTTACGG TTCTGGCCT TTGCTGGCC TTTGCTCAC AAAACGAGGTG TACAAGAAAG GACGAAATAG CGTTGGCCG GAAAATGCC AAGGACCGGA AAACGACCGG AAAACGAGGTG TACAAGAAAG GACGAAATAG |
| 4621 | CCCTGATTCT GTGGATAACC GTATTACCGC CTTTGAGTGA GCTGATACCG CTCGGCCAG CGAACGAC GGGACTAAGA CACCTATTGG CATAATGGCC GAAACTCACT CGACTATGGC GAGGGGGTC GGCTTGCTGG |
| 4691 | GAGGCCAGCG AGTCAGTGA CGAGGAAGCG GAAGAGGCC CAATAACGCC ACCGCCTCTC CCCGGCGTT CTCGCGTGCCTCACTC GCTCTCGC CTTCCTCGC GTTATCGGT TGGGGAGAG GGGCGGCCAA |
| 4761 | GGCCGATTCA TTAATGCAGC TGGCACGCCA GGTTCCCGA CTGGAAAGGC GGCAGTGGAGC GCAACGCAAT CCGGCTAAGT AATTACGTCTG ACCGTGCTGT CCAAAAGGCT GACCTTCTGC CGGTCACTCG CGTTGGCGTTA |
| 4831 | TAATGTTAGT TAGCTCACTC ATTAGGCACCC CCAAGGCTTTA CACTTTATGC TTCGGCTCG TATGTTGTGT ATTACACTCA ATCGAGTGA G TAATCCGTGG GTTCCGAAAT GTGAAATACG AAGGCCGAGC ATACAACACA |

4901 GGAATTGTGA GCGGATAACA ATTTCACACA GGAAACAGCT ATGACCATGA TTACGCCAG CGGCCAATTAA
CCTTAACACT CGCCTATTGT TAAAGTGTGT CCTTTGTCGA TACTGGTACT AATGCGGTTC GCGCGTTAAT

~~~~~  
BssHII

~~~~~

NcoI

~~~~~  
KpnI

~~~~~

4971 ACCCTCACTA AAGGGAAACAA AAGCTGGGTA C
TGGGAGGTGAT TTCCCTTGTGTT TTCGACCCAT G

FIG._48I

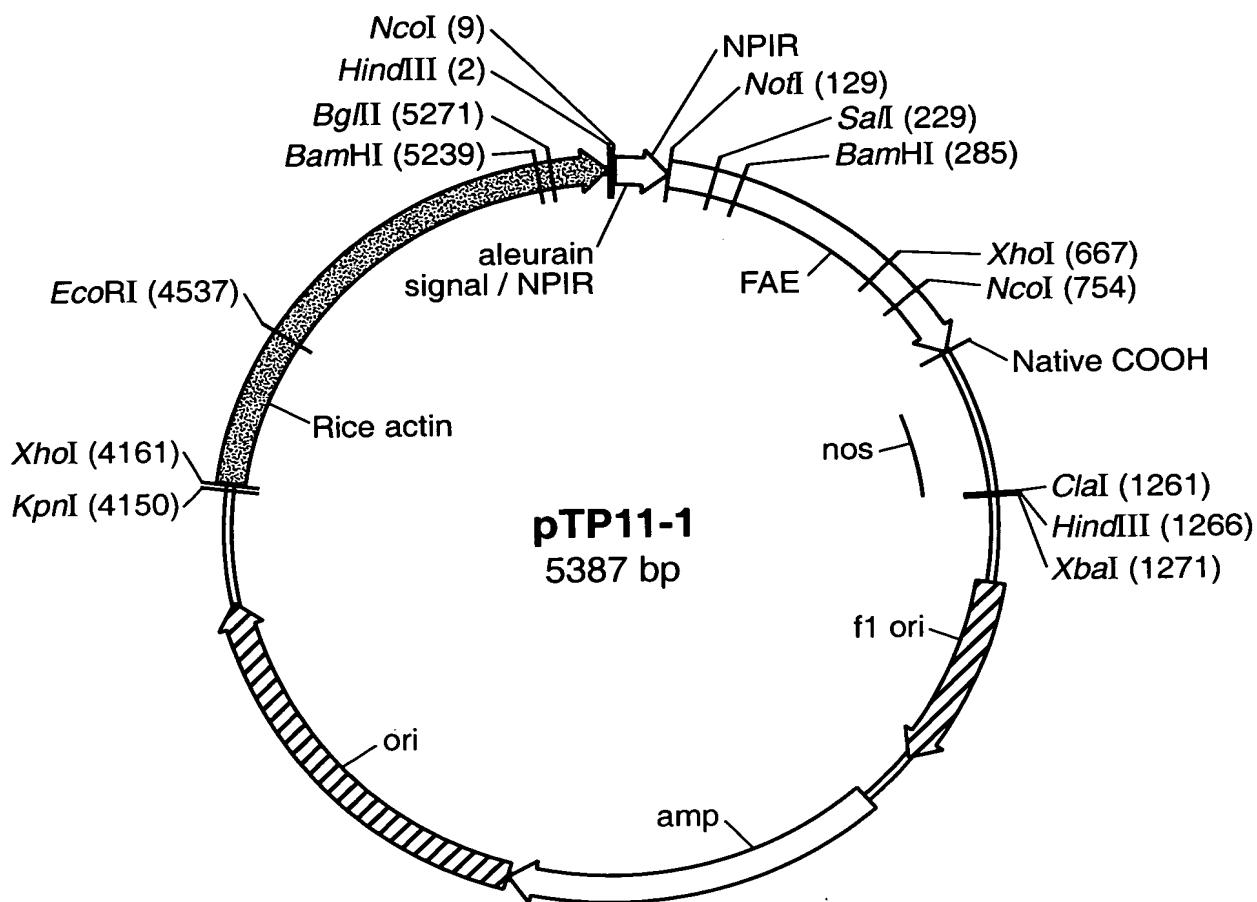


FIG._49A

147 / 154

NcoI

HindIII

~~~~~  
M A H A R V L L A L A V L A T A A V A V  
1 AAGCTTACCA TGGCCACGC CGCGTCCCTC CTCCTGGGCC TCGCCGTGCT GCCCACGGCC GCCGTGCCG

Not I

~~~~~  
• A S S S F A D S N P I R P V T D R A A S T
71 TCGCCTCCTC CTCCCTCCTC GCCGACTCCA ACCCGATCCG GCCCGTCAACC GACCGCGGG CGGCCTCCAC
• Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAAGACC TCTACAGGCC TTAGTCGAA ATGGCCACTA TCTCCAAAGC TGCCTACGCC

SalI

AccI

~~~~~  
D L C N I P S T I I K G E K I Y N S Q T D I N G  
211 GACCTGTGCA ACATTCGGTC GACTATTATC AAGGGAGAGA AAATTACAA TTCTCAAACT GACATTAACG  
BamHI

~~~~~  
• W I L R D D S S K E I I T V F R G T G S D T N
281 GATGGATCCT CGCGGACGAC AGCAGCAAG AAATAATCAC CGTCTTCCGT GGCACGTGGTA GTGATAAGAA
• L Q L D T N Y T L T P F D T L P Q C N G C E V
351 TCTACAACTC GATACTAATC ACACCCCTCAC GCCTTTCGAC ACCCTAACAC AATGCAACGG TTGTGAAGTA
H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGGTA
• Q Y P D Y A L T V T G H X L G A S L A A L T A
491 GCCAGTATCC GGAATACGGG CTGACCGTG CCGGCCAAC CCTCGGGCC TCCCTGGGG CACTCACTGC
• A Q L S A T Y D N I R L Y T F G E P R S G N Q
561 CGCCCAAGCTG TCTGGGACAT ACGACAAACAT CGGCTGTAC ACCTTCGGCG AACCGGGCAG CGGCAATCAG

XbaI

~~~~~  
A F A S Y M N D A F Q A S S P D T T Q Y F R V T  
631 GCCTTCGGGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAAC GACGGCAGTAT TTCCGGGTCA

FIG.-49B

NcoI

|      |                                                                                      |       |
|------|--------------------------------------------------------------------------------------|-------|
| 701  | H A N D G I P N L P P V E Q G Y A H G G V E Y .                                      |       |
|      | CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCCATGGG GTGCTAGAGTA        |       |
| 771  | W S V D P Y S A Q N T F V C T G D E V Q C C E                                        |       |
|      | CTGGAGCGTT GATCCTTACA GCGGCCAGAA CACATTTGTC TGCACTGGG ATGAAGTGC A GTGCTGTGAG         |       |
| 841  | A Q G G Q G V N N A H T T Y F G M T S G A C T W                                      |       |
|      | GCCCCAGGGCG GACAGGGTGT GAATAATGCC CACACGACTT ATTTGGGAT GACGAGCGGA GCCTGTACAT         |       |
|      | *                                                                                    |       |
| 911  | GGTGATCAGT CATTTCAGG TCCCGAGTG TACCAAGAAA GATGGATGTC CTTGGAGAGGG GGCAGGGTAA          |       |
| 981  | CCACTGAGG ATGAGCTGTA AAGAAGCGA TCGTTCAAAAT ATTTGGCAAT AAAGTTCTT AAGATTGAAAT          |       |
| 1051 | CCTGTTGCCG GTCTTGCGAT GATTATCATA TAATTCTGT TAAGGATCTGT TAAGGATCTGT ATAATTAAACA       |       |
| 1121 | TGTAATGCAT GACGTTATT ATGAGATGGG TTTTTATGAT TAGAGTCCCCG CAATTATACCA TTTAATACGC        |       |
|      | Clai                                                                                 |       |
| 1191 | GATAGAAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGGGGG TGTCACTAT GTTACTAGAT             |       |
|      | HindIII                                                                              |       |
|      | ~~~~~                                                                                |       |
|      | Clai                                                                                 | XbaI  |
|      | ~~~~~                                                                                | ~~~~~ |
| 1261 | CGATAAGCTT CTAGAGGGC CGGTGGAGCT CCAATTGCC CTATAGTGAG TCGTATTACG CGCGCTCACT           |       |
| 1331 | GGCCGGTCTT TTAACACGTC GTGACTGGGA AAACCCCTGGC GTTACCCAAC TTAATCGCC TGCAGGCACAT        |       |
| 1401 | CCCCCTTTCG CCAGCTGGG TAATAGGAA GAGGGCCGC CCGATGCC CTCACACAG TTGCGCAGGCC              |       |
| 1471 | TGAATGGCGA ATGGGACGG CCCTGTAGCG GGCATTAAG CGGCATTAAG CGGGGGGGT GTGGTGGTTA CGCGCAGGGT |       |
| 1541 | GACCGCTACA CT"GCCAGCG CCCTAGGCC CGCTCCTTC GCTTCTTCC CTTCCCTTCT CGCCACGTT             |       |
| 1611 | GCCGGCTTC CCCGTCAAGC TCTAAATCGG GGGCTCCCTT TAGGGTCCCG ATTAGTGTCT TTACGGCACC          |       |
| 1681 | TCGACCCAA AAAACTTGAT TAGGGTGTAG GTTCACGTAG TGGGCCATCG CCCTGTATAGA CGGTTTTTCG         |       |
| 1751 | CCCTTTGACG TTGGAGCTCA CGTCTTTAA TAGTGGACTC TGTTCACAAA CTGGAACAAAC ACTCAACCCCT        |       |
| 1821 | ATCTCGGTCT ATTCTTTTGA TTATAAGG ATTTGGGA TTTGGCCCA TTTCGGCTA TTGGTTAAAA AATGAGCTGA    |       |
| 1891 | TTAACAAAA ATTAAACGGA AATTTAAC AAATTTAAC GCTTACATT TAGGTGGCAC TTTCGGGA                |       |
| 1961 | AATGTGGCG GAACCCCTAT TTGGTTATT TTCTAAATAC ATTCAAAAT GTATCCGCTC ATGAGACAAAT           |       |
| 2031 | AACCCCTGATA ATGCTTCAA TAATATTGA AAAGGAAGAG TATGAGTATT CAACATTCC GTGTCGCCCT           |       |
| 2101 | TATTCCTTT TTGCGGGCAT TTGCTTTC TGTGGCTTGC TGGTTTTGCT CACCCAGAAA CGCTGGTAA AGTAAAGAT   |       |
| 2171 | GCTGAAGATC AGTGGGGTGC ACGAGTGGGT TACATCGAAC TGGATCTCAA CAGGGGTAAAG ATCCTTGAGA        |       |
| 2241 | GTTTTGCCG CCAAGAACGT TTTCCTAATGA TGAGCACTT TAAAGTTCTG CTATGTGGCG CGGTATTATC          |       |
| 2311 | CCGTATGAC GCCGGGCAAG AGCAACTCGG TGGCCGCATA CACTATCTC AGAATGACTT GGTTGAGTAC           |       |
| 2381 | TCACCAAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCCAGTGC GGCATAACCA        |       |

FIG.-49C

2451 TGAGTGATAA CACTGGGGCC AACTTACTTC TGACAACGT CGGAGGACCG AGGAGCTAA CCCGCTTTTT  
2521 GCACAAACATG GGGGATCATG TAACTCGCCT TGATCGTGG GAAACGGAGC TGAATGAAGC CATAACAAAC  
2591 GACGAGCGTG ACACCAACGAT GCCTGTAGCA ATGGCAACAA CGTTGGCAA ACTATTAACG GGCAGAACTAC  
2661 TTACTCTAGC TTCCCGGCAA CAATTAAATAG ACTGGATGGA GGGGGATAAA GTTGGAGGAC CACTTCTGCG  
2731 CTCGGCCCTT CCGGCTGGCT GGTTTATTGC TGATAAATCT GGAGGCCGTT AGCGTGGGTC TCGGGGTATC  
2801 ATTGCAGCAC TGGGGCCAGA TGGTAAGGCC TCCCAGTATCG TAGTTATCTA CACGACGGGG AGTCAGGCCAA  
2871 CTATGGATGA ACCAAATAGA CAGATCGCTG AGATAGGTGC CTCACITGATT AAGCATTGGT AACTGTCAAGA  
2941 CCAAGTTAC TCATAATAC TTAGATTGA TTAAAGACTT CATTTTAAAT TTAAAGGAT CTAGGTGAAG  
3011 ATCCCTTTTG ATAATCTCAT GACCAAAATC CTTAACGTG AGTTTTCTGTT CCACTGAGCG TCAAGACCCCG  
3081 TAGAAAAGAT CAAAGGATCT TCTTGAGATC CTTTTTTCT GGGCGTAATC TGCTGCTTGC AAACAAAAA  
3151 ACCACCGCTA CCAGGGGGGG TTGTTTGCGC GGATCAAAGAG CTACCAACTC TTGTTCCGAA GGTAACTGGC  
3221 TTCAAGAG CGCAGATAACC AAATACTGTC CTTCTAGTGT AGCCGTAGTT AGGCCACAC TTCAAGAAACT  
3291 CTGTAGACC GCCTACATAC CTCGGCTCTGC TAATCCTGTT ACCAGTGGCT GCTGCCAGTG GCGATAAAGTC  
3361 GTGTCTTAC GGGTGGACT CAAGACGATA GTTACCGGAT AAGGGCGAGC GGTGGGGCTG AACGGGGGGT  
3431 TCGTGCACAC AGCCCAAGCTT GGAGGAACGG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG  
3501 AAAGGCCAC GCTTCCCGAA GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTGC GACAGGAGA  
3571 GGCACGAGG GAGCTTCCAG GGGAAAAGC CTGGTATCTT TATAGTCCTG TCGGGTTCTG CCACCTCTG  
3641 CTTGAGGTC GATTTTGTG ATGCTCGTCA GGGGGCGGA GCCTATGGAA AAACGCCAGC AACGGGGCT  
3711 TTTACGGTT CCTGGCCTT TGCTGGCCTT TTGCTCACAT GTTCTTCCCT GCGTTATCCC CTGATTCTGT  
3781 GGATAACCGT ATTACCGCT TTGAGTGGAGC TGATACCGCT CGCGCAGCC GAACGACCGA GGCAGCGAG  
3851 TCACTGAGCG AGGAAGGGGA AGAGGGCCCA ATACGCAAAC CGCCCTCTCCC CGGGCGTGG CGGATTCAATT  
3921 AATGCAGCTG GCAGGACAGG TTTCGGACT GAAAGGGGG CAGTGGCGC AACGCAATTAA ATGTTGAGTTA  
3991 GCTCACTCAT TAGGCACCCC AGGCTTACA CTTATGCTT CGGCTCGTA TGTGTTGGG AATTGTTGAGC  
4061 GGATAACAAT TTACACAGG AACAGCTAT GACCATGATT Xhol KpnI

4131 GGGAAACAAA GCTGGGTACC GGGCCCCC TCGAGGTCT TCATATGCT GAGAAGAGAG TCGGGATACT  
4201 CCAAAATAAA ACAAAAGTAA GATTAACCTG TCAAAAGTGA AAACATCACT TAAAGGTGG TATAAGTAA  
4271 ATATCGTAA TAAAAGTGG CCCAAAGTGA AATTACTCT TTCTACTAT TATAAAATT GAGGATGTT  
4341 TGTGGTACT TTGATACGTC ATTTTGTAT GAATTGGTT TAAAGTTAT TCGGGATTTG GAATGCGATA  
4411 TCTGTATTTG AGTCGGTTT TAAGTTCGTT GCTTTGTAATACAGAGGG ATTGTGTTAA GAATATCTT  
~~~~~ EcorI  
~~~~~  
4481 TAAAAAACCC ATATGCTAAT TTGACATAAT TTGAGAGAAA AATATATATT CAGGGCAATT CCACAATGAA  
4551 CAATAATAAG ATTTAAATAG CTTGGCCCG TTGAGCGAT GGGTATTGTT TCTAGTAAA TAAAAGATAAA  
4621 ACTTAGACTC AAAACATTAA CAAAAACAC CCCTAAAGTC CTAAGGCCA AAGTGGCTATG CACGATCCAT

FIG.\_49D

150 / 154

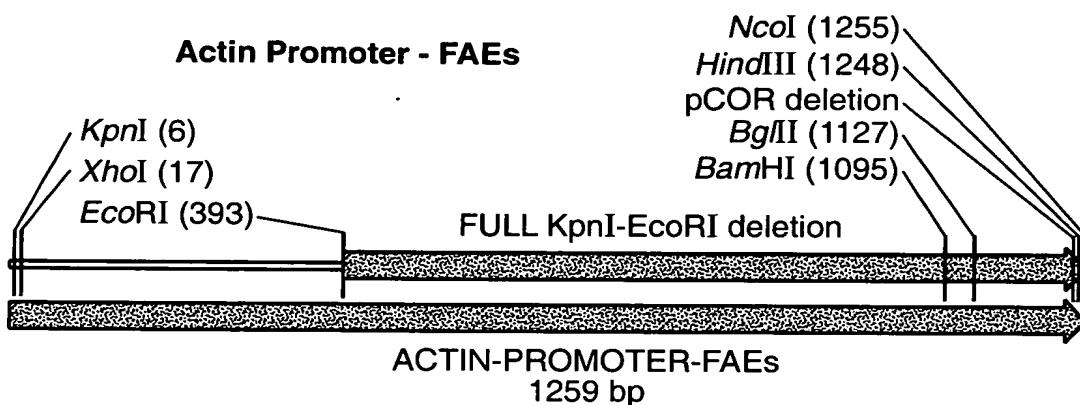
4691 AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCCACCCCA GTGCAGCCAA CTGGCAATA GTCTCCACCC  
4761 CCGGCACAT CACCGTGAAT TGTCCGCAC ACCGCACGTC TCGCAGCCAA AAAAAGAAAA AGAAAGAAAA  
4831 AAAAGAAAAA GAAAACAGC AGGTGGTTC AGGAGGAGG GGGTGTGGG GGGGGAAA GCGAGGGAA TCGCGAGCAG  
4901 CGACGAGGCC CGGGCCCTCC TCCGCTTCCA AAGAAACGCC CCCCCATCGCC ACTATATACA TACCCCCCCC  
4971 TCTCCTCCCA TCCCCCAAC CCTAACCA CCACCAAC CACCTCCTCC CCCCTCGCTG CCGGACGACG  
5041 AGCTCCCTCCC CCCTCCCCCT CGGGCCGC CGGTAACAC CCCGGCCCTC TCCCTCTTCT TTCTCCGTT  
5111 TTTTTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT GGGTGGGGA GAGGGGCTC GTGGCCAGA  
BamHI ~~~~~~

5181 TCGGTGGCG GAGGGGGGG GATCTCGCGG CTGGGTCTC CGGGCGTGA G TCGGCCGGGA TCCTCTGGGG  
BglII ~~~~~~

5251 GAATGGGGCT CTCGGATGTA GATCTTCTT CTTTCTTCTT TTTGTGGTAG ATTGTGAAATC CCTCAGCATT  
5321 GTTCATCGGT AGTTTTCTT TTCAATGATT GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGC

**FIG.-49E**

151 / 154



|     | KpnI                                                                                                                                                          | XhoI |  |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------|------|--|
| 1   | <u>GGTACCGGGC</u> <u>CCCCCTCGA</u> <u>GGTCATTCA</u> <u>ATGTTGAGA</u> <u>AGAGAGTCGG</u> <u>GATAGTCAA</u> <u>AATAAAACAA</u>                                     |      |  |
|     | <u>CCATGGCCCG</u> <u>GGGGGGAGCT</u> <u>CCAGTAAGTA</u> <u>TACGAACCTCT</u> <u>TCTCTCAGCC</u> <u>CTATCAGGTT</u> <u>TTATTTGTT</u>                                 |      |  |
| 71  | <u>AGGTAAGATT</u> <u>ACCTGGTCAA</u> <u>AAGTGAAAC</u> <u>ATCAGTTAAA</u> <u>AGGTGGTATA</u> <u>AGTAAAATAT</u> <u>CGGTAATAAA</u>                                  |      |  |
|     | <u>TCCATTCTAA</u> <u>TGGACCAAGTT</u> <u>TTCACTTTTG</u> <u>TAGTCAATT</u> <u>TCCACCATAT</u> <u>TCATTTATA</u> <u>GCCATTATTT</u>                                  |      |  |
| 141 | <u>AGGTGGCCCA</u> <u>AAGTGAAATT</u> <u>TACTCTTTTC</u> <u>TACTATTATA</u> <u>AAAATTGAGG</u> <u>ATGTTTGTC</u> <u>GGTACTTTGA</u>                                  |      |  |
|     | <u>TCCACCGGGT</u> <u>TTCACTTTAA</u> <u>ATGAGAAAAG</u> <u>ATGATAATAT</u> <u>TTTAACCTCC</u> <u>TACAAAACAG</u> <u>CCATGAAACT</u>                                 |      |  |
| 211 | <u>TACGTCA</u> <u>TTT</u> <u>TTGATGAAT</u> <u>TGGTTTTAA</u> <u>GTTTATTCGC</u> <u>GATTTGGAAA</u> <u>TGCATATCTG</u> <u>TATTTGAGTC</u>                           |      |  |
|     | <u>ATGCAGTAA</u> <u>AACATACTTA</u> <u>ACCAAAATT</u> <u>CAAATAAGCG</u> <u>CTAACCTTT</u> <u>ACGTATAGAC</u> <u>ATAAAACTCAG</u>                                   |      |  |
| 281 | <u>GGTTTTAAG</u> <u>TTCGTTGCTT</u> <u>TTGTAAATAC</u> <u>AGAGGGATT</u> <u>GTATAAGAAA</u> <u>TATCTTAAA</u> <u>AAACCCATAT</u>                                    |      |  |
|     | <u>CCAAAAATT</u> <u>AAGCAACGAA</u> <u>AACATTATG</u> <u>TCTCCCTAAA</u> <u>CATATTCTT</u> <u>ATAGAAATT</u> <u>TTTGGGTATA</u>                                     |      |  |
|     | <u>EcoRI</u>                                                                                                                                                  |      |  |
| 351 | <u>GCTAATTG</u> <u>CATAATT</u> <u>TTT</u> <u>GAGAAAATA</u> <u>TATATTCA</u> <u>GG</u> <u>CGAATTCCAC</u> <u>AATGAACA</u> <u>AT</u> <u>AATAAGATTA</u>            |      |  |
|     | <u>CGATTAAACT</u> <u>GTATTAAAAA</u> <u>CTCTTTTAT</u> <u>ATATAAGTCC</u> <u>GCTTAAGGTG</u> <u>TTACTTGT</u> <u>TTATTCTAAT</u>                                    |      |  |
| 421 | <u>AAATAGCTT</u> <u>CCCCCGTTGC</u> <u>AGCGATGGGT</u> <u>ATTTTTCTA</u> <u>GTAAAATAAA</u> <u>AGATAAAACTT</u> <u>AGACTCAAAA</u>                                  |      |  |
|     | <u>TTTATCGAAC</u> <u>GGGGGCAACG</u> <u>TCGCTACCCA</u> <u>TA</u> <u>AAAAGAT</u> <u>CATT</u> <u>TTTATT</u> <u>TCTATTGAA</u> <u>TCTGAGTTT</u>                    |      |  |
| 491 | <u>CATT</u> <u>ACAAA</u> <u>AACAACCC</u> <u>CTAAAGTCTAA</u> <u>AGCCCAAAGT</u> <u>GCTATGCACG</u> <u>ATCCATAGCA</u> <u>AGCCCAGCCC</u>                           |      |  |
|     | <u>GTAAATGTT</u> <u>TTGTTGGGA</u> <u>TTTCAGGATT</u> <u>TCGGGTTTCA</u> <u>CGATACGTG</u> <u>TAGGTATCGT</u> <u>TCGGGTGCGG</u>                                    |      |  |
| 561 | <u>AACCCAA</u> <u>CCC</u> <u>AACCCAA</u> <u>CCC</u> <u>ACCC</u> <u>CAGTGC</u> <u>AGCCA</u> <u>ACTGG</u> <u>CAAATAGTCT</u> <u>CCACCCCCGG</u> <u>CACTATCACC</u> |      |  |
|     | <u>TTGGGTTGGG</u> <u>TTGGGTTGGG</u> <u>TGGGTCACG</u> <u>TCGGTTGACC</u> <u>GTTTATCAGA</u> <u>GGTGGGGCC</u> <u>GTGATAGTGG</u>                                   |      |  |
| 631 | <u>GTGAGTTGTC</u> <u>CGCACCA</u> <u>CCG</u> <u>CACGTCTCGC</u> <u>AGCC</u> <u>AAAAAA</u> <u>AAAAAAAGAA</u> <u>AG</u> <u>AAAAAA</u> <u>GA</u> <u>AAAAGAAA</u>   |      |  |
|     | <u>CACTCAACAG</u> <u>GCGTGGTGGC</u> <u>GTGCA</u> <u>GAGCG</u> <u>TCGGTTTTT</u> <u>TTTTTTCTT</u> <u>TCTTTTTT</u> <u>CTTTTTCTT</u>                              |      |  |
| 701 | <u>AACAGCAGGT</u> <u>GGGTCCGGGT</u> <u>CGTGGGGGCC</u> <u>GGAAAAGCGA</u> <u>GGAGGATCGC</u> <u>GAGCAGCGAC</u> <u>GAGGCCCGGC</u>                                 |      |  |
|     | <u>TTGTCGTCCA</u> <u>CCCAGGCCA</u> <u>GCACCCCCGG</u> <u>CCTTTCGCT</u> <u>CCTCCTAGCG</u> <u>CTCGTCGCTG</u> <u>CTCCGGGCCG</u>                                   |      |  |

**FIG.\_50A**

152 / 154

771 CCTCCCTCCG CTTCCAAAGA AACGCCCCC ATGCCACTA TATACATACC CCCCCCTCTC CTCCCATCCC  
GGAGGGAGGC GAAGGTTCT TTGCGGGGGG TAGCGGTGAT ATATGTATGG GGGGGGAGAG GAGGGTAGGG  
  
841 CCCAACCTA CCACCACCCAC CACCACCAAC TCCTCCCCC TCGCTGCCGG ACGACGAGCT CCTCCCCCT  
GGGTTGGGAT GGTGGTGGTG GTGGTGGTGG AGGAGGGGGG AGCGACGGCC TGCTGCTCGA GGAGGGGGGA  
  
911 CCCCCCTCCGC CGCCGCCGGT AACCAACCCG CCCCTCTCCT CTTCTTTCT CCGTTTTTT TTTCGTCTCG  
GGGGGAGGCG GCGGCCGCCA TTGGTGGGCG GGGGAGAGGA GAAAGAAAGA GGCAAAAAAA AAAGCAGAGC  
  
981 GTCTCGATCT TTGGCCTTGG TAGTTGGGT GGGCGAGAGC GGCTTCGTG CCCAGATCGG TGCGCGGGAG  
CAGAGCTAGA AACCGGAACC ATCAAACCA CCCGCTCTCG CCGAAGCAGC GGGTCTAGCC ACAGCGCCCTC

BamHI

~~~~~

1051 GGGCGGGATC TCGCGGCTGG CGTCTCCGGG CGTGAGTCGG CCCGGATCCT CGCGGGGAAT GGGGCTCTCG
CCCGCCCTAG AGCGCCGACC GCAGAGGCC GCACTCAGCC GGGCCTAGGA GCGCCCTTA CCCCAGAGAGC

BglII

~~~~~

1121 GATGTAGATC TTCTTCCTT CTTCTTTTG TGGTAGAATT TGAATCCCTC AGCATTGTTA ATCGGTAGTT  
CTACATCTAG AAGAAAGAAA GAAGAAAAAC ACCATCTAA ACTTAGGGAG TCGTAACAAG TAGCCATCAA

HindIII NcoI

~~~~~ ~~~~~

1191 TTTCTTTCA TGATTGTGA CAAATGCAGC CTCGTGCGGA GCTTTTTGT **AGGTAGAAGC** TTACCATGG
AAAGAAAAGT ACTAAACACT GTTTACGTG GAGCACGCCT CGAAAAAACCA **TCCATCTCG** AATGGTACC

KpnI-EcoRI - deletion underlined and restored NCO site in bold in vectors pJQ4.9,
pJQ3.2 and pJO6.3.

FIG._50B

153 / 154

ALEURAIN_d I t d NPIR (Ap plast) Structur and Sequ nc



ALEURAIN-NPIR-DEL
93 bp

+1 M A H A R V L L L A L A V L A T A A V A

HindIII NcoI

~~~~~ ~~~~~~

1 AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCCG  
TTCGAATGGT ACCGGGTGCG GGGCCAGGAG GAGGACCGCG AGCGGCACGA CCGGTGCCGG CGGCAGCGGC

+1 V A S S R A A

NotI

~~~~~

71 TCGCCTCCTC CCGCGCGGCC GCC
AGCGGAGGAG GGCAGCGCCGG CGG

FIG._51

SEE1 (Senescence enhanced) PROMOTER sequence

1 CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
PstI XbaI XbaI
~~~~~  
71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGTTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC  
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT  
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACCGCGCTC CCAACAAAAT ATCGTCCCCC  
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTCGCGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC  
SmaI  
~~~~~  
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
SphI
~~~~~  
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTGG CTTAATTGAC TTTATTTTG  
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCAACCC ATCTGAATAT CGAGACGGAT AATAGGCTGG  
631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTCT GGGCATTACA GCTGGAGGCT  
PstI  
~~~~~  
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCGGCACC GGGACGCGAG CTCCCGCTA CCAGTACCAT CTCGCCTCGC TCCCCCTGCC GGACGACCCA
841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATG
ATG

FIG._52

SEE1 (Senescence enhanced) PROMOTER plus vacuolar aleurain SIGNAL/NPIR sequence

1 CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
PstI XbaI XbaI
~~~~~  
71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGTTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC  
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT  
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACCGCGCTC CCAACAAAAT ATCGTCCCCC  
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTCGCGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC  
SmaI  
~~~~~  
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
SphI
~~~~~  
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTGG CTTAATTGAC TTTATTTTG  
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCAACCC ATCTGAATAT CGAGACGGAT AATAGGCTGG  
631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTCT GGGCATTACA GCTGGAGGCT  
PstI  
~~~~~  
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCGGCACC GGGACGCGAG CTCCCGCTA CCAGTACCAT CTCGCCTCGC TCCCCCTGCC GGACGACCCA
841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATG
ATG
M A H G R I L F L A L A V L
BssHII
NotI
· A T A A V A A A S L A D S N P I R P V T E R A ·
911 TGGCCACCGC CGCGGTGGCC GCCGCATCNT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGCGCGC
NotI
~~~~~  
· A A  
981 GGCGGCC  
FIG.\_53